

# A Review on Data Analytics for Supply Chain Management: A Case study

**Anitha P**

Department of Information Science and Engineering  
JSS Academy of Technical Education, Bengaluru-560060, India  
Email: anitha.palakshappa@gmail.com

**Malini M. Patil**

Department of Information Science and Engineering  
JSS Academy of Technical Education, Bengaluru-560060, India  
Email: drmalinimpatil@gmail.com

Received: 25 February 2018; Accepted: 24 May 2018; Published: 08 September 2018

**Abstract**—The present study bridges the gap between the two intersecting domains, data science and supply chain management. The data can be analyzed for inventory management, forecasting and prediction, which is in the form of reports, queries and forecasts. Because of the price, weather patterns, economic volatility and complex nature of business, the forecasts may not be accurate. This has resulted in the growth of Supply chain analytics. It is the application of qualitative and quantitative methods to solve relevant problems and to predict the outcomes by considering quality of data. The issues like increased collaboration between companies, customers, retailers and governmental organizations, companies are adopting Big Data solutions. Big Data applications can be linked for Supply Chain Management across the fields like procurement, transportation, warehouse operations, marketing and also for smart logistics. As supply chain networks becoming vast, more complex and driven by demands for more exacting service levels, the type of data that is managed and analyzed also becomes more complex. The present work aims at providing an overview of adoption of capabilities of Data Analytics as part of a “next generation” architecture by developing a linear regression model on a sales-data. The paper also covers the survey of how big data techniques can be used for storage, processing, managing, interpretation and visualization of data in the field of Supply chain.

**Index Terms**—Big Data, Supply Chain Management, Supply Chain Analytics, Supply Chain Network, regression analysis, Smart logistics, Big Data Management Systems.

## I. INTRODUCTION

After the 1990s great changes in the operating rules of world economy and market competition patterns, enterprises have identified the need for globalization of economic development. Companies need to rely on the integration of their own and external resources available

in the market. This integration enables the working pattern of a system that gives quick response to the needs of the market. Also the ad-hoc situations that arise in the market. Such a system is referred to as Supply Chain Management (SCM) [1]. SCM is defined as managing the flow of information, material and resources across and within the network of upstream and downstream organizations [2]. Supply chain can be defined as network of flow of products, financial deliverables, customer services and network of information, material and resources. Management of multiple relationships in supply chain is referred to as supply chain management. Some of the factors like product success, customer satisfaction, growth of an organization depends on the successful execution of SCM [5].

The term Big Data has been used initially by two NASA researchers during 1997 to refer to the visualization challenge for systems with large amount of data sets which are ubiquitous in nature [6]. Big data can be understood as the data which is complex, large in volume, rapid growing with numerous, autonomous and independent data sources [33]. Data has increased in various fields on a large scale basis from couple of decades, because of which the term “Big Data” has been coined [20]. Big data has its impact in various domains: helps in renovating the supply chain, manages the customer fidelity in marketing, health, optimizing the route and reducing cost in transportation, reducing the risk in finance etc. [7]. Deployment of Big data management system for Supply chain management will achieve greater benefits as system becomes more agile.

Big Data can be defined as a large volume of data – both unstructured and structured. Due to rise in the field of social media, Internet of things and mobile devices, it is found that there is a massive increase in real time data generation [1]. As per the survey, more than 1200 Exabyte’s of data are generated every year from different data sources [1]. Most of the data generated are not structured. Amount of unstructured data is approximately 80%, where these data are difficult to store, analyze and process.

The analysis of Big Data leads to insights that help in taking better decisions and strategic business moves which is termed as Analytics.

Complete process of extraction of Big Data consists of two things- data management and data analytics. Data management comprises “processes and supporting technologies to gather data, store data, to prepare and retrieve it for analysis. The techniques used to analyze and gather intelligence from Big data is referred to as Big Data

Analytics. Analytics is defined as the collection and analysis of data in terms of qualitative and quantitative for decision making. BDA is the application of advanced analytic techniques to huge data sets. Gartner explains that only 15% of Fortune 500 companies will be able to make full use of big data for creating value and only 8% of them are currently using Big Data Analytics [1].

Big data analytics has preceded its need for the Supply chain management of any organization and many companies are struggling to unveil its business value [2]. Big data analytics challenge is to analyze the irregular patterns of data arriving next to the present huge data sets [26]. BDA is important in creating an integrated view of operational performance and customer satisfaction of both sender and recipient in the SCM [16]. It is really challenging to meet fully the Supply chain susceptibility because of the complexity in Supply chain components, process, supplier and services [25]. Supply chain analytics has been developed its own identity in supply chain management by using the Business Intelligence tools for the analysis of customer behavior, optimization of upstream and downstream operations and also insight on advanced routing solutions. Examples are component

complexity, supplier complexity, process complexity and service complexity [3].

According to Waller and Fawcett (2013), research area in the field of Big Data intersecting with SCM could illuminate a “great number of new opportunities” for both academia and practitioners [2]. They also pointed out that very sparse literature survey about predictive analytics, data science and Big Data are available in the field of SCM. SCM and Logistics are not the new ideas. Logistics can be defined as the process of managing the procurement, flow of products, storage of materials, parts and finished inventory to maximize the profit through the cost effective fulfillment of orders [10]. SCM is an extensive field than logistics. The concept of logistics evolved as a subfield of Supply chain management whose main vision is creating & presenting a single plan of flow of materials and information. During 1980 logistics has been defined as to fulfill taking care, transportation, loading and unloading, packing and processing between the manufacturer and consumer for commodity and various other functions [28]. SCM superimposes the logistics framework in implementing the connectivity and coordination between entities like suppliers and customers [10].

According to the professionalism, skills, flexibility, reliability, attitude, behavior, reputation and integrity of human resources in logistics companies are mainly important from the client’s point of view [23]. Fig.1 shows the flow of supply chain management. It explains various activities involved in manufacturing a product and services starting from planning till delivery of materials to end customer. Truly speaking, SCM is not a chain of processes instead it is a network of multiple businesses and relationships [8].

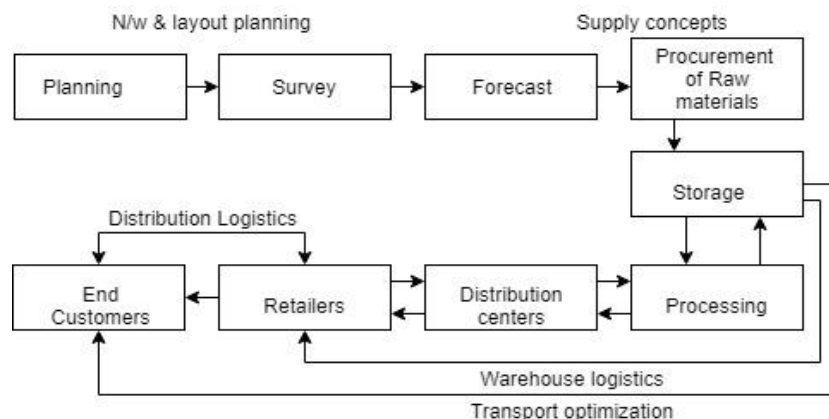


Fig.1. Process of Supply Chain Management

## II. BIG DATA

The name Big Data has been used first by two NASA researchers in 1997 as a challenge for visualizing large data sets. There after researchers and specialists in the field of Information Management have been gradually paying attention towards Big Data. The process for Big Data is classified into 7 phases as shown in the Fig 2. Data Discovery phase includes accessing the resources

needed for solving the Big data analytics and one should familiarize with the data [18]. In data preparation stage, ETL (Extract, Load, Transform, Load) operations are applied on the required data. Huge volumes of data from different sources causes high probabilities of errors [30]. So data needs to be transformed, cleaned and audited before they are loaded into Data warehouses [30]. Technologies can be used if required in this phase. In the subsequent phases, especially in the next phase project team has to decide the usage of methods, workflows and

techniques required for analyzing the data along with evaluation and interpretation of data. In the model execution phase, the model chosen in the previous phase is executed with appropriate data sets available. Once the results are available, communication of the results and optimizing the results if possible has been carried out in the remaining phases [18].

Big Data has a positive impact in different domains: helps in reconditioning the supply chain, increasing sales and marketing, real route optimization etc. [7]. Big data analytics process has been explained from the perspective of supply chain data [18]. The following section provides the taxonomy of ten main attributes of Big data applications in Supply Chain management.

#### 1. Volume:

Volume refers to the huge amount of data generated from emails, twitter, photos, and videos every second. In SCM, Volume can be related to the data generated from the use of Sensors, bar codes, ERP, Transport management system and database technologies. Previously volume is measured in Gigabytes which is now measured in Zettabytes (ZB) or even Yottabytes (YB). There are different forms and ways of storing the Big data generated from supply chain industry [20]. Rational database management system (RDBMS), which is a structured model employed to see, analyze, manufacture and store the huge amount of supply chain management data. Also the data clusters in Big data storage includes components like

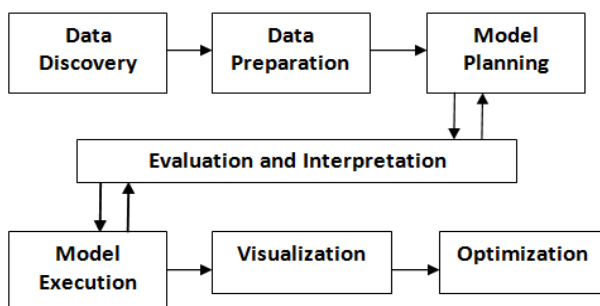


Fig.2. Big Data Process.

- Direct attached storage (DAS) - includes different types of hard disks/ hard drives which are attached to DBMS
- Network storage (NS) – which comes in two forms Network attached storage(NAS) and Storage Area Network(SAN) [20]

#### 2. Velocity:

It mainly refers to the speed of the data collected, analyzed and transferred. It impacts on the efficiency and decision making models and algorithms in the field of SCM.

#### 3. Variety:

It refers to the different forms of data like structured, unstructured or semi structured [35]. Also it includes different types of data from XML to video to SMS. The

variety of data in the field of SCM includes data from diverse sources like retailers, distributors, suppliers, inventory, sales, consumer [9] etc. Big Data collection process in SCM includes 2 variety of sources: Upstream and downstream sources. The data from upstream source includes supplier's side, through intermediate stream or warehouse side. Data from downstream includes logistics, distribution or retailer side [20].

#### 4. Veracity:

Correctness or trustworthiness of data is referred to as veracity. This verifies the quality of data from SCM, compliance issues etc.

#### 5. Value:

It refers to monetary worth of data. It is challenging to monitor the value of reports, statistics, impacts on the insights etc.

#### 6. Variability:

Lack of variability in big data can be defined as the data which is not consistent or liable to vary or change. Supply Chain variability in terms of information sharing, integration, quality control, unexpected delays in the supply process etc.

#### 7. Visualization:

Analyzing the data graphically is termed as visualization. Visualization method is more effective in conveying meaning than spreadsheets and reports or using numbers and formulas in terms of Supply chain data can be visualized using ERP, custom developed reports or using graphical method.

#### 8. Virality:

It measures speed of data movement from one network to other. From the supply chain management view, it is very essential for logistics process to be carried out.

#### 9. Viscosity:

It mainly refers to the data latency or the delay in data. It can be easily understood as an element of velocity.

#### 10. Volatility:

How long the data is valid and how long it should be stored. It is mainly associated with old and new data.

**RESEARCH FINDING 1:** From literature survey it is found that Supply chain management is a big network of multiple business strategies and relationships. In the Big data ecosystem, where the data is found to be completely ubiquitous, it is challenging to justify the Big data dimensions (all V's). For example, the Big data dimension volume relates to the data generated from Transport management system, Enterprise resource planning and many more. Similar kind of explanation can be found in another dimensions of Big data that is Variety, which is more often referenced under data collection process in SCM. Other issues related to Big data dimensions are quality of data, information sharing, development of customized report, logistics process,

validation of data and others are really challenging to achieve the completeness of all dimensions of Big data for SCM.

### III. DATA ANALYTICS AND TECHNIQUES USED IN SUPPLY CHAIN MANAGEMENT

In spite of the largest growth in the field of data analytics experienced by customer insight, Analytics has many applications across end to end supply chain. As the acquisition and transportation cost per entry is driven to be minimum, there is an inevitably corrupted measurements and errors in the large scale data has been found [29]. Since any of the data sources continuously generate data in real time, analytics must often be performed [29]. Applications of advanced analytic techniques has been described to supply chain management [14]. Supply Chain Data Analytics has been classified into three types of Analytics: Descriptive, Predictive and Prescriptive analytics.

#### A. Descriptive Analytics:

Descriptive Analytics (DA) is mainly used to analyze “what is happening” now in order to answer the question of “What happened” in the past. This is the first level of analytics where 90% of organizations apply this strategy for betterment of the future. DA identifies the historical data and analyzes the pattern. Descriptive Analytics mainly aims at identifying the problems and opportunities in the field of SCM within the existing processes and functions [17].

Descriptive Analytics uses the techniques like

- Data Modeling
- Regression Analysis
- Visualization
- OLAP (online analytical processing) operations like drill down, up and across to identify the areas.

Big data tools for supply chain analytics has been summarized in the Table 1. OLAP operations for Supply chain may include shipments, products, logistics, customers, suppliers and other dimensions like rates and cost. The applications of Descriptive analytics provide the managers with real time data regarding the quantities of goods and location in the supply chain.

#### B. Predictive Analytics:

Predictive analytics (PA) use both quantitative and qualitative methods to analyze the real time and historical data to estimate the past and future levels of integration of business processes among functions or companies, as well as the associated costs and service levels [9]. Predictive analytics aims at projecting what will happen in the future and why it may happen [17]. PA includes algorithms/techniques such as [2]

1. Time series methods & Advanced forecasting. These methods are used for predicting the sales in SCM.
2. Statistical algorithms such as Discriminant Analysis, k-NN, Naive Bayes (NB) and Bayes Networks (BN).
3. Decision trees, CART and Random Forests uses the hierarchical sequential structure
4. Clustering algorithms used to group homogeneous elements in a data set.
5. Frequent pattern mining algorithms

Predictive analytics mainly focused on forecasting at strategic, tactic and operational levels, which is based on the planning process in terms of network design, production planning, inventory management and capacity planning [14]. Predictive analytics uses mathematical algorithms and programming in order to predict the patterns within data

To understand the Descriptive and Predictive Analytics, an experimental is performed on sales data, which is a benchmark dataset available. The results are found to be interesting. A predictive model is developed based on regression analysis.

Detailed explanation is provided in section VI.

#### C. Prescriptive Analytics:

DA and PA are focused on what and when it will happen, whereas Prescriptive Analytics anticipates on “why it has happened”. It collects the data continuously to re-predict the events which enable the decision makers to increase the prediction accuracy for taking better decisions. Prescriptive analytics explains the reasons behind certain events. It is mainly associated with simulation and optimization [2]. The aim of Prescriptive analytics is to improve the business performance [17]. Three classes of algorithms used under this analytics method are

- Decision trees
- Fuzzy Rule-Based System
- Switching Neural Networks (Logic Learning Machine)

Prescriptive analytics is focused on the optimization of mathematical and simulation techniques in order to provide the decision support tools which has been built on descriptive and predictive methods.

RESEARCH FINDING 2: Big data tools available for supply chain analytics mainly used for data exploration, integration of data, statistical analysis, proper visualization methods and understanding the data warehouse system. Few of them are R-prog, Informatica, PASW. The main observation is about LINGO, DSM, which are mainly used for documentation purpose based on customer support system. Pentaho is used to handle structured and unstructured volumes of data. To

summarize the integration of all tools is an important task in SCA for developing a Decision support system.

Table 1. Big data tools for Supply Chain Analytics

Name of the Tool	Description
LINGO	Optimization tool for linear, nonlinear and mathematical programming which is introduced by John H Thomson in 1989. This tool can be used for Easy model expressions, convenient data options and also it helps for documentation
DSM	Drop shipping management tool which is mainly used to increase arbitrage opportunities. This is a best tool in the field of drop shipping arbitrage. It can be used for handling customer support based on the sales history, ticketing system and statistical reports.
R programming	It is an object oriented, free, open source software environmental tool for statistical analysis and graphics. Introduced by “Ross Ihakaa and Robert Gentleman”. It mainly based on command line interfaces. It is used by statisticians and data miners which supports for statistical tests, linear and nonlinear modeling and also for time series analysis.
Mongo DB	It is a document oriented database which replaces the concept of a “row” with a more flexible model. Dwight Merriman, Eliot Horowitz and Kevin Ryan. Mongo DB supports for 1. Indexing 2. Aggregation 3. Special collection types and 4. File storage
Pentaho	Pentaho serves as an analytical/software platform for Big Data [20]. It generates reports from both structured and unstructured volumes of data. It provides services for Businessmen for easy access, visualization, integration and data exploration[20].
Informatica	Informatica introduces a series of products relates to data integration and warehousing. It extracts data from heterogeneous data sources, combine and standardizes the data which will be presented in a uniform format for the purpose of processing.
PASW	This is used to perform data analysis collected from different sources and presentation functions which includes graphical and statistical analysis. Some of the features used statistical data analysis are descriptive statistics and sophisticated inferential and multivariate statistical procedures.
Hadoop - Map reduce	It is a method used to process vast amounts of data and also a model based on java for distributed computing. Map-reduce algorithm contains two main tasks- Map and Reduce. This technique splits the input data sets into independent chunks which are processed by the map task in a parallel way, sort the data and given as an input for the reduce task. This framework concentrates on scheduling the tasks, monitoring the scheduled task and also re-executing the tasks which will fail during execution.

IV. CHALLENGES OF SUPPLY CHAIN MANAGEMENT

According to Robak et al [2014], the open research problem in supply chain management along with logistics can be analyzed from the view of stake holders and executive business components, where key business

functions are forecasting, inventory management, transport management and also human resources. Big data can address the issues of Supply chain like timely response, time delivery, real time planning, supplier and customer relationship management etc. [15]. Operating an effective supply chain involves continuous flow of information, which in turn helps to create better material flow [8]. The main focus in supply chain is the customer. So achieving a good customer focused system is one of the aim of SCM [8]. The key supply chain processes/ challenges are listed and shown in the Fig 3.

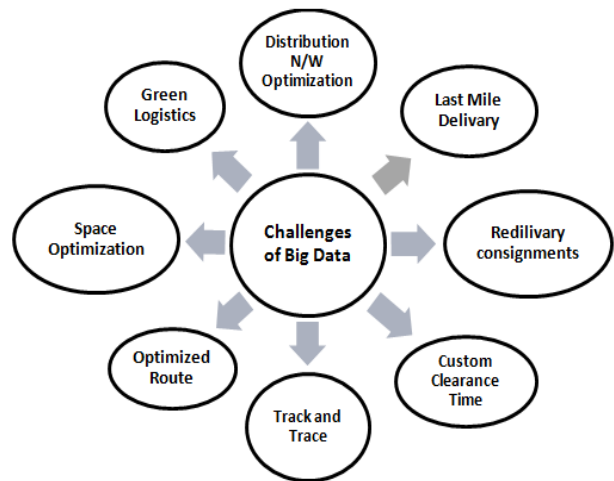


Fig.3. Big Data Challenges in Logistics

Supply chain need to be aware of the benefits given by the big data for their operations.

- A. *Distribution Network optimization*
- B. *Green Logistics*
- C. *Route Optimization*
- D. *Space Optimization*
- E. *Last mile delivery problem*
- F. *Redelivery Consignments*
- G. *Custom clearance time*
- H. *Track & Trace*

A. *Distribution Network Optimization*

Distribution network involves locating warehouses and production plants and also to identify the strategy for distribution of products from supplier to warehouses and from warehouses to customers [12]. The key challenges of Distribution network optimization are optimization of domestic transportation network, achieve flexibility in network redeployment when network changes and also to minimize the total cost of distribution network. Benefits of n/w optimization by achieving above challenges are the Optimized Network.

B. *Green Logistics*

It implies efficiency in using transportation equipment’s. Some of the key features in Green logistics are Minimization of carbon emission, Cost reduction from resource saving and environmental externalities [13]. Benefits by overcoming the mentioned challenges are

better utilization of resources, social and environment responsibility etc. Introducing or adopting green logistics is a complex process which requires cross disciplinary coordination and also changes in the current operation process [19]. This can be also being achieved by introducing new practices in the area of supply and distribution that links them to other participants like suppliers and customers in the value chain. This link must be supported by management staff's, their characteristics and also by human resources [19].

### C. *Route Optimization*

Route optimization is the very important factor to control efficiently the physical flow of supply chain. Some of the challenges are like Optimization of single route trucking trip, Allocate appropriate resources per trip, Cost reduction. By achieving these goals, the benefits from Route optimization is Route Efficiency.

### D. *Space Optimization*

The key parameters in Space optimization are Maximizing space utilization, improve productivity and also to minimize the cost. Benefits from Space Optimization is utilization of space in a better way.

### E. *Last mile delivery problem*

The main challenge in this problem is delivering thousands of packages to customer in an efficient way. Another challenge here is time bound delivery of goods. Benefit by overcoming this problem is Customer Satisfaction. BDA enable the Last mile delivery problem by increasing the level of operational efficiency [15].

### F. *Redelivery Consignments*

Some of the parameters in Redelivery consignments are proper packaging and handling, Efficient transportation which reduces the redelivery of the products. Benefits by achieving above parameters is efficiency in monetary through minimizing the redelivery.

### G. *Custom clearance time*

The main parameter in custom clearance time is to maintain proper documentation which represents the client in the time of custom examination and assessment. The benefit by achieving the given parameter is to avoid detention charges.

### H. *Track and Trace*

Some of the parameters under track and trace is "Near-real-time "tracking and Status & position information [11]. Advantage by achieving this parameter is Performance improvement in track & trace.

RESEARCH FINDING 3: Literature survey reveals that the Big data challenges in Logistic management of supply chain system mainly relates to the stake holders that is customers and the key Business functions. It is found that network optimization, route optimization and space optimization form the basis for a proper management of physical space, physical flow of Supply chain and proper strategy of distribution of products to

customers. Minimization of cost, time and space are the main parameters. The above challenges are also reflecting the development of customer end activities such as efficient delivery of products, packaging and handling and documentation processes of related activities. A complete model can be developed by taking a real world data and establish that the above challenges are met using Big data approach

## V. SUSTAINABLE SMART LOGISTICS

Logistics is a part of supply chain management. Logistics can be defined as process of managing the procurement, storage and movement of goods along with the related information flow to maximize the profit of the organization through cost effective fulfillment of orders. Logistics has been identified as a core element of supply chain management [21]. The aim of the Logistics is to serve the customers in a cost effective way. Sustainability in logistics can be defined as a cultural issue based on the demonstration of many companies and organizations. It can be a trend setting the business model or setting up the new market opportunities and also preparing for future scenarios [22]. The term intelligent or smart logistics can be defined as a different logistics operation which are planned, managed and controlled in a smart way compared to conservative solutions [21]. Besides the planning, managing & controlling the objects and resources of logistics, also the aggregation and processing of the collected data is an important task of Smart logistics [27]. Some of the approaches to improve logistics by making them more intelligent are as follows

### A. *Autonomous Logistics*

It describes the ability of logistics objects to process the information, to provide and to execute their own decisions.

### B. *Product intelligence*

The way of storing and transporting any physical order or product instance in an efficient manner.

### C. *Intelligent transport systems*

It mainly refers to the innovative services related to transport and traffic management. This enables the user to be better informed, safer and smarter use of transport network.

### D. *Physical Internet*

Physical internet suggests exploiting the digital internet metaphor in order to develop a physical internet towards meeting the global logistics sustainability challenge.

### E. *Intelligent cargo*

Capabilities under Intelligent Cargo are self-identification, context detection, access to services, status monitoring and registering.

### F. *Self-organizing logistics*



Self-organized logistic company's functions without intervention by managers, engineers or by the software control. Of the above approaches, the Intelligent traffic management systems attracted the greatest interest during recent years. Uckelmann described Smart Logistics as technical components to gain data on the level of material flow and to process these data for monitoring and further purposes [23]. Smart Logistics are a key approach in the information logistics cross company and international transportation networks to meet the need for robustness, flexibility, resilience and agility [23]. Technical components in smart logistics are as follows:

- A. RFID for Identification
- B. RTLS like GPS and others for location
- C. Usage of Sensors.

RFID is used to ensure a secure identification of the different objects at all stages of supply chain. For example, in the use case of vehicle development process up to 50 single items are placed in one vehicle, which needs to be securely identified. Smart logistics using RFID can be used to identify objects placed in the vehicle. One of the application of RFID is to record the ID and sensed temperature given by the sensors during transportation [34].

## VI. REGRESSION

Regression is a collection of data mining or statistical modelling technique used to describe the behavior of random variable by using one or more quantitative variables. The Linear regression uses the method of a straight line ( $y = mx + c$ ) in order to predict the value of  $y$  by determining the appropriate values for  $m$  and  $c$  based upon the value of given  $x$ .  $y$  becomes a dependent variable and other variables becomes a predictor or independent variable. The relationships between target variable and the predictors are summarized in model. This relationship has been applied to variety of datasets whose target values are unknown. Some of the applications of regression analysis in Business are predictive analytics, operation efficiency, supporting decisions, correcting errors and new insights. Regression analysis is mainly used to determine the values of parameters for a function which cause the function to best fit of asset data that we provide. Equation mentioned below explains the relationships in terms of symbols. Equation explains that regression is a function of estimating the value of a continuous target expressed as  $Y$ , a function ( $F$ ) of one or more predictors ( $x_1, x_2, \dots, x_n$ ), a set of parameters ( $\theta_1, \theta_2 \dots \theta_n$ ) and a measure of error as ( $e$ ) [32].

$$Y = F(x, \theta) + e. \quad (1)$$

The target variable can be understood as dependent variables or response variables and predictors as independent or explanatory variables. Parameters of regression are also known as regression coefficients [32].

### A. Linear regression

Linear regression is a statistical method which is used to learn about relationship between response variables and a predictor variable. Regression analysis is a best choice when all of the independent variables are continuous valued. Straight line linear regression analysis is the simplest form of regression which contains a single predictor variable  $y$  and models as a linear function of  $x$ .

$$Y = mx + c. \quad (2)$$

### B. Coefficients of determination

The coefficient of determination denoted by  $R$  is an output of regression analysis. This has been interpreted as follows

- The  $R$  ranges from 0 to 1.
- $R$  is 0 means the dependent variable cannot be predicted from an independent variable.
- If  $R$  is between 0 to 1 means that the extent to which the dependent variable is predictable [32].
- $R$  is 0.10 indicates that 10% of variance in  $y$  is predictable from  $x$ , and if it is 0.20 then 20% is predictable and so on.

## VII. EXPERIMENTS & RESULTS

This section comprises of discussion on experiments, results, data set description, summary of scatter plot representation and about PSPP.

### A. Data Set Description:

The sample sales data set size taken for regression analysis is 25. The data set is in .csv format. The attributes of sales-data and their description is provided in the Table 2.

Table 2. Data set description

Attribute	Description
Store number	Number for each stores
Sales_quantity_Feb	Quantity of sales in Feb
Sales_Feb	Amount of sales happened in Feb
GP	Gross Profit
Expected_Sales	Result of sales after performing linear regression

### B. PSPP:

PSPP is an open source software which can be used for the analysis of sampled data. Originally developed in late 1990s, intended as a free alternative for IBM SPSS Statistics. It has a graphical user interface and conventional command-line interface. It is written in C and uses GNU Scientific Library for its mathematical routines. PSPP is used for statistical analysis of sampled datasets. This tool reads the data, analyzes according to the commands and writes the result to output window. The language is similar to SPSS statistical products. The important features of PSPP are frequencies, cross-tabs

comparison of means (t-tests and one-way ANOVA); linear regression, logistic regression and many more.

C. Scatter plot:

In order to determine the linear relationship between the variables (dependent and independent), it is suggested to run a scatter plot on the given data set. If the graph contains no linear relationship, then no need for linear regression. From Fig 4, it is found that points on the graph are linear. This indicates that linear relationship exists between the variables and simple regression can be applied. The scatter plot obtained is shown in Fig.4

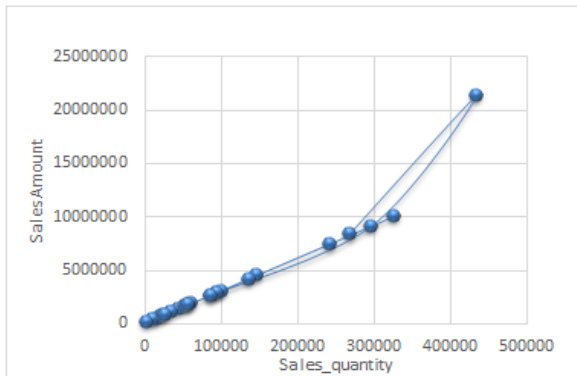


Fig.4. Scatter plot to check linearity

D. Interpretation of Results:

The results obtained from regression model run on PSPP are depicted in Table 3(a-c) and Table 4. The tabulations are related to model summary, ANOVA, Coefficients. Table 4 depicts the results generated from simple linear regression along with a scatter plot in Fig 5. According to the table, predictions are almost nearing to the actual values. It is approximately 80%.

The interpretation of coefficient of determination R obtained from model summary is in the range of 0 & 1, that is 0.96. This indicates that the dependent variable can be predicted. In the present experiment the dependent variable is sales\_quan\_Feb and shown in the Table 3(a).

Analysis of variance (ANOVA) is the collection of statistical models and their related procedures. It is used to analyze the differences among the group means. It is

found from the table that the significance of ANOVA for sales\_quan\_Feb is found to be 0.

The regression coefficients for the predictor (Expected\_sales) depicts the difference in response per unit difference in the predictor. They are tabulated in Table 3(c).

From the coefficients Table 3(c) regression equation is shown below.

$$DV=23411.25+0.02 * IV \tag{3}$$

Where

DV = Dependent variable

IV = Independent variable

Table 3. (a): Model Summary

Model Summary (sales\_quan\_feb)

R	R Square	Adjusted R Square	Std.Error of the Estimate
.96	.92	.92	32816.24

Table 3. (b): ANOVA

ANOVA (sales\_quan\_feb)

	Sum of Squares	df	Mean Square	F	Sig.
Regression	282643336179.97	1	282643336179.97	262.46	.000
Residual	23691920426.65	22	1076905473.94		
Total	306335256606.63	23			

Table 3. (c) : Coefficients

Coefficients (sales\_quan\_feb)

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std.Error	Beta		
(Constant)	23411.25	8709.12	.00	2.69	.013
Sales_feb	.02	.00	.96	16.20	.000

Table 4. Simple Regression Table

Sls_Qty_Jan	Sales_Jan	GP_Jan	Sls_Qty_Feb	Sales_Feb	GP-Feb	Expected_Sales
255205	8504700	1573382	267965	8349929	1652051	7573852
413922	12475303	1831377	434618	21322029	1922946	12284188
282774	9839422	1849295	296913	9108444	1941759	8392048
81951	2387691	417905	86048	2639719	438801	2432089
91253	2819919	539118	95816	2939359	566074	2708175
43466	1660907	193122	45640	1400098	202778	1289984
231654	6353325	984933	243236	7461811	1034180	6874903
49835	1065829	507583	52326	1605223	532962	1478960
95128	4148480	589751	99884	3064161	619238	2823155
89868	3755131	703207	94361	2894734	738368	2667051



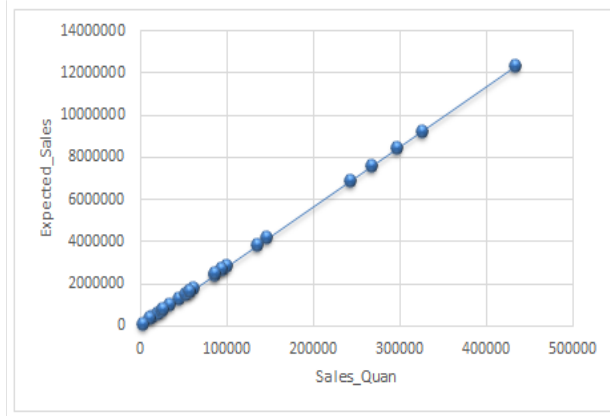


Fig.5. Scatter plot after prediction.

### VIII. CONCLUSION

The bibliographic survey conducted in this paper mainly focuses on the Supply chain management activities related to data warehouse, marketing and transportation. The Business transformation and improvements in operation cost have contributed to the development of the Supply chain analytics. A predictive model developed on sample sales-data using linear regression method is found to be 80% nearer to the available results. The model also provides information about the descriptive statistics through which the descriptive analytics can be done. Further the work is enhanced using multiple and polynomial regression methods. The present work provides a good platform for the Big data analytics and possible integration of BDA and SC. It can act as a good customer support system.

### ACKNOWLEDGEMENT

The authors extend their gratitude to their organization JSS Academy of Technical Education, Bengaluru, for the facilities provided in order to carry out the research work.

### REFERENCES

- [1] Feki, Mondher, Imed Boughzala, and Samuel Fosso Wamba. "Big Data Analytics-Enabled Supply Chain Transformation: A Literature Review." 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 2016
- [2] Rozados, Ivan Varela, and Benny Tjahjono. "Big data analytics in supply chain management: Trends and related research." 6th International Conference on Operations and Supply Chain Management. 2014.
- [3] Kao, Gio, et al. "Supply chain lifecycle decision analytics." Security Technology (ICCST), 2014 International Carnahan Conference on. IEEE, 2014.
- [4] Mishra, Deepa, et al. "Big Data and supply chain management: a review and bibliometric analysis." *Annals of Operations Research* (2016): 1-24.
- [5] Kamble, Shridhar, Aaditya Desai, and Priya Vartak. "Data mining and data warehousing for supply chain management." *Communication, Information & Computing Technology (ICICT)*, 2015 International Conference on. IEEE, 2015.
- [6] Cox, Michael, and David Ellsworth. "Application-controlled demand paging for out-of-core visualization." *Proceedings of the 8th conference on Visualization'97*. IEEE Computer Society Press, 1997.
- [7] Benabdellah, Abla Chaoui, Et Al. "Big Data for Supply Chain Management: Opportunities and Challenges."
- [8] Lambert, Douglas M., and Martha C. Cooper. "Issues in supply chain management." *Industrial marketing management* 29.1 (2000): 65-83.
- [9] Waller, Matthew A., and Stanley E. Fawcett. "Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management." *Journal of Business Logistics* 34.2 (2013): 77-84.
- [10] Christopher, Martin. *Logistics & supply chain management*. Pearson UK, 2016.
- [11] Jakobs K., Pils C., Wallbaum M. (2001) Using the Internet in Transport Logistics - The Example of a Track & Trace System. In: Lorenz P. (eds) *Networking — ICN 2001*. Lecture Notes in Computer Science, vol 2093. Springer, Berlin, Heidelberg
- [12] Amiri, Ali. "Designing a distribution network in a supply chain system: Formulation and efficient solution procedure." *European Journal of Operational Research* 171.2 (2006): 567-576.
- [13] Dekker, Rommert, Jacqueline Bloemhof, and Ioannis Mallidis. "Operations Research for green logistics—An overview of aspects, issues, contributions and challenges." *European Journal of Operational Research* 219.3 (2012): 671-679.
- [14] Souza, Gilvan C. "Supply chain analytics." *Business Horizons* 57.5 (2014): 595-605.
- [15] Robak, S., Franczyk, B., Robak, M. (2014). Research Problems Associated with Big Data Utilization in Logistics and Supply Chain Design and Management. *Annals of Computer Science and Information Systems*, 3(1), 245-249.
- [16] Mikavicaa, Branka, Aleksandra Kostić- Ljubisavljevića, and Vesna Radonjić. "Big data: challenges and opportunities in logistics systems." 2nd Logistics Intl. Conference. 2015.
- [17] Wang, Gang, et al. "Big data analytics in logistics and supply chain management: Certain investigations for research and applications." *International Journal of Production Economics* 176 (2016): 98-110.
- [18] Robak, Silva, Bogdan Franczyk, and Marcin Robak. "Business process optimization with big data analytics under consideration of privacy." *Computer Science and Information Systems (FedCSIS)*, 2016 Federated Conference on. IEEE, 2016. Hypothesis
- [19] Seroka-Stolka, Oksana. "Green Initiatives in Environmental Management of Logistics Companies." *Transportation Research Procedia* 16 (2016): 483-489.
- [20] Addo-Tenkorang, Richard, and Petri T. Helo. "Big data applications in operations/supply-chain management: A literature review." *Computers & Industrial Engineering* 101 (2016): 528-543.
- [21] McFarlane, Duncan, Vaggelis Giannikas, and Wenrong Lu. "Intelligent logistics: Involving the customer." *Computers in Industry* 81 (2016): 105-115.
- [22] Ceniga, Pavel, and Viera Sukalova. "Future of logistics management in the process of globalization." *Procedia Economics and Finance* 26 (2015): 160-166.
- [23] Palšaitis, Ramūnas, Kristina Čižiūnienė, and Kristina Vaičiūtė. "Improvement of Warehouse Operations Management by Considering Competencies of Human Resources." *Procedia Engineering* 187 (2017): 604-613.

- [24] Uckelmann, Dieter. "A definition approach to smart logistics." International Conference on Next Generation Wired/Wireless Networking. Springer Berlin Heidelberg, 2008.
- [25] Kao, G., Lin, H., Eames, B., Haas, J., Fisher, A., Michalski, J., ... & Wyss, G. (2014, October). "Supply chain lifecycle decision analytics". In Security Technology (ICCST), 2014 International Carnahan Conference on (pp. 1-7). IEEE
- [26] Leveling, J., Edelbrock, M., & Otto, B. (2014, December). Big data analytics for supply chain management. In Industrial Engineering and Engineering Management (IEEM), 2014 IEEE International Conference on (pp. 918-922). IEEE.
- [27] Kirch, Martin, Olaf Poenicke, and Klaus Richter. "RFID in Logistics and Production—Applications, Research and Visions for Smart Logistics Zones." *Procedia Engineering* 178 (2017): 526-533.
- [28] Jian-qiang, Wu, Zhang Lei, and Zhu Guo-qing. "Performance-based Evaluation on the Logistics Warehouse." *Procedia Engineering* 11 (2011): 522-528.
- [29] Slavakis, Konstantinos, Georgios B. Giannakis, and Gonzalo Mateos. "Modeling and optimization for big data analytics: (statistical) learning tools for our era of data deluge." *IEEE Signal Processing Magazine* 31.5 (2014): 18-31.
- [30] Abai, Nur Hani Zulkifli, Jamaiah H. Yahaya, and Aziz Deraman. "User requirement analysis in data warehouse design: a review." *Procedia Technology* 11 (2013): 801-806.
- [31] Park, Sung H. "Simple linear regression." *International Encyclopedia of Statistical Science*. Springer Berlin Heidelberg, 2011. 1327-1328.
- [32] Srimani P.K., Patil M.M. (2014) Regression Model for Edu-data in Technical Education System: A Linear Approach. In: ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India- Vol II. Advances in Intelligent Systems and Computing, vol 249. Springer, Cham.
- [33] Nivedita Das, Leena Das, Siddharth Swarup Rautaray, Manjusha Pandey, " Big Data Analytics for Medical Applications", *International Journal of Modern Education and Computer Science(IJMECS)*, Vol.10, No.2, pp. 35-42, 2018.DOI: 10.5815/ijmeecs.2018.02.04.
- [34] Ping-Ho, Ting. "An efficient and guaranteed cold-chain logistics for temperature-sensitive foods: applications of RFID and sensor networks." *International Journal of Information Engineering and Electronic Business* 5.6 (2013):
- [35] Pradeep Kumar M. Kanaujia, Manjusha Pandey, Siddharth Swarup Rautaray, "A Framework for Development of Recommender System for Financial Data Analysis", *International Journal of Information Engineering and Electronic Business(IJIEEB)*, Vol.9, No.5, pp. 18-27, 2017. DOI: 10.5815/ijieeb.2017.05.03

## Authors' Profiles



**Dr. Malini M. Patil** is presently working as an Associate Professor in the Department of Information Science and Engineering at J.S.S. Academy of Technical Education, Bangalore, Karnataka, INDIA. She received her Ph.D. degree from Bharathiar University in the year 2015. Her research interests are big data analytics, bioinformatics, cloud computing, image processing. She has published more than 20 research papers in many reputed international journals and guiding four students. She has attended and presented papers in many international conferences in India and Abroad. Published article, entitled "Performance analysis of Hoeffding trees in data streams by using massive online analysis framework" in *International Journal of Data Mining, Modelling and Management*, Inderscience Publishers. Published article, entitled "Mining Data streams with concept drift in massive online analysis frame work", *WSES Transactions on computers*", She is a member of IEEE, ISTE, CSI, IEI. She is a recipient of Distinguished Woman in Science Award for the year 2017 from Venus International Foundation. She has received a best paper presenter award in Second International Conference of Data Management, Analytics and Innovation – ICDMAI-2018, Pune, India. The award was sponsored by springer. Contact email: drmalinimpatil@gmail.com



**P. Anitha** is currently working as an Assistant Professor in the Department of Information Science and Engineering at JSSATE, Bengaluru. She is a research scholar in the field of Supply Chain Data Analytics at JSSATE Research Centre, Department of CSE, affiliated to Visveswaraya Technological university, Belagavi, India under the guidance of Dr. Malini M Patil, Associate Professor, Department of ISE, JSSATE, Bengaluru. She has completed her M. Tech in 2012 from Department of Computer science, BMSCE, Bengaluru under the Visveswaraya Technological University. She is a life member of Indian Society for Technical Education.

**How to cite this paper:** Anitha P, Malini M. Patil, " A Review on Data Analytics for Supply Chain Management: A Case study", *International Journal of Information Engineering and Electronic Business(IJIEEB)*, Vol.10, No.5, pp. 30-39, 2018. DOI: 10.5815/ijieeb.2018.05.05