

An Approach for Forecast Prediction in Data Analytics Field by Tableau Software

Bibhudutta Jena

Email: bibhudutta93@gmail.com

Received: 13 December 2017; Accepted: 18 October 2018; Published: 08 January 2019

Abstract—The current era is generally treated as the era of data, Users of computer are gradually increasing day by day and vast amount of data is generated from multiple domains such as healthcare- domain, Business related domains etc. The terminology Business Intelligence (BI) generally refers different technologies, applications and practices used for the collection, integration, analysis, and presentation of information of business related domain. The main motive for Business intelligence and analytics are to help in decision making process and to enhance the profit of the organisation. Various business related tools are used to analyze & visualize different types of data which are generated frequently. Tableau prepared its mark on the Field of BI by being one of the first companies to permit business customers the ability to achieve equitably arduous data visualization in a very interesting, drag and drop manner. Tableau will enhance decision making, add operational awareness, and increase performance throughout the organization The presented paper describes different tools used for business intelligence field and provides a depth knowledge regarding the tableau tool. It also describes why tableau is widely used for data visualization purpose in different organization day by day. The main aim of this paper is to describe how easily forecasting and analysis can be done by using this tool ,this paper has explained how easily prediction can be done through tableau by taking the dataset of a superstore and predict the forthcoming sales and profit for the next four quarters of the forthcoming year. In the collected dataset sales and profit details of different categories of goods are given and by using the forecasting method in tableau platform these two measures are calculated for the forthcoming year and represented in a fruitful way. Finally, the paper has compared all the framework used for business intelligence and analytics on the basis of various parameters such as complexity, speed etc.

Index Terms—Big Data, Business Intelligence, Data Visualization, Tableau, QlickView.

I. INTRODUCTION

Due to the increased market competition increased data management and analysis has landed as in an era that requires further optimization data management and analysis. Big data technologies like apache HADOOP

provide a frame work for parallel data processing and generation of analyzed results .MAPREDUCE method is used for analysis of data using various data analysis algorithms like clustering, fragmentation and aggregation. (BI) generally refers different technologies, applications and practices used for the collection, integration, analysis, and presentation of information of business related domain[1]. The main motive for Business intelligence and analytics are to help in decision making process and to enhance the profit of the organisation. Various business related tools are used to analyze & visualize different types of data which are generated frequently. Data is one of the most important and vital aspect of different activities in today's world. Now -a -days the modern civilization gradually more dependent on the age of information which is concentrated in computers, So the Society is becoming more dependent on computers and technology for functioning in every- day life .As the number of users are gradually increasing day by day, So the data generated by them is also enlarging gradually[2].The main aim of this paper is to describe how easily forecasting and analysis can be done by using this tool ,this paper has explained how easily prediction can be done through tableau by taking the dataset of a superstore and predict the forthcoming sales and profit for the next four quarters of the forthcoming year. The reminder of this paper is organized as. Section II presents a literature survey on various approaches and tools used for business intelligence and analysis purpose. Section III presents a brief idea about Tableau tools and its application. The motivation for this work is laid out in Section II. Section IV represents the Implementation part for the forecast analysis of sample retail store data Finally, Section V concludes the paper.

II. LITERATURE SURVEY

Tableau presently has the largest customer domain. When businesses organization are focusing for a BI solution, it's always their first choice. In the below section the paper has described different BI tools that are currently used in the business market for enhancing the business strategy. Sisense is one of the leaders in the market of BI and a winner of the Best Business Intelligence Software Award given by Finances Online [3][4], It was considered as one of the most adopted business, software review platforms. The users of this

tool will also enjoy effectively by using the in-chip technology in a database that operates the data much more faster than the existing traditional systems. BIRT project is a flexible, Standalone, and 100% pure Java reporting tool for creating and presenting reports in opposition to data sources varying from typical business relational databases, to XML data sources, to in-memory Java objects. It is originated as a top-level project integrating with the Eclipse Foundation. Similarly icCube is a SaaS end to end BI platform, particularly to be ingrained in to the developed application[5][6]. It amalgamates smoothly with any application due to on-the-fly-authentication and authorization ability. icCube can be treated as a dream for any software developer who wants to present predicated dashboards or a solid web based self service BI solution, to their end-users. It is not Cost effective. Business Optimization Software Domos imports together the customers, the data, and the acumen business users demand to present a detailed view of what's happening in the organization[7][8]. The Jaspersoft Business Intelligence Suite provides different ways for end customers to initiate interactive analysis. In the case of casual user, this might include simply changing a filter setting on a report to present various slice of data. This tool is generally helpful for data analyst to writing powerful, multi-dimensional expressions. Microsoft also provides Business Intelligence platform comprises of Analysis Services, Integration Services, Master Data Services, Reporting Services, and several client applications used for initiating or operating with analytical data. But this software is not able to make the dashboard attractive like other BI tools. This data is often travelled across various databases in multiple locations with different versions of database software. Pentaho platform provides an easier way for preparing and blending any data and adds a spectrum of tools to easily analyze, visualize, explore, report and predict[9]. In the above section the paper has discussed some business intelligence tools which are earlier used for the business intelligence and business analysis processes. In the next section the paper has described how tableau is helpful for business intelligence and analysis purpose which plays an important role for decision making processes in the organisation.

III. TOOL STUDY

Tableau prepared its mark on the Field of BI by being one of the first companies to permit business customers the ability to achieve equitably arduous data visualization in a very interesting, drag and drop manner. Tableau's data virtualization is head and shoulders above what traditional BI vendors offer[10]. Tableau occupied its mark on the world of BI by being one of the first companies to give business customers the ability to operate fairly complex data visualization in a very intuitive, drag and drop manner. In this platform visualization and analysis can be done without having proper knowledge in programming domain, The output analyzed results can be shown in an effective manner. In

the below section the paper has described the how smoothly the forecasting process can be done by tableau tool by taking a business test case. Tableau clearly and beautifully visualizes your data. Tableau has the functionality to present reports on supreme large sets of data without enormously affecting network performance. someone can easily go from novice to expert level with proper assistance through this tool.

IV. IMPLEMENTATION

This section describes how the data is processed in tableau platform and how forecasting can be done with the collected dataset. The data which are collected for processing are from a sample store present in USA which comprises of different attributes or fields , In the below figure describes the excel format of data that is taken for processing in tableau platform. There are multiple ways to import the data set in tableau software, One of the biggest leverage of tableau software is, it allows the user to import the data in portable document format and perform different data analytic operation by utilizing the data. Tableau's data virtualization is leader and shoulders above what conventional BI vendors offer. Tableau done its mark on the world of BI by being one of the first companies to give business users the propensity to perform fairly complex data visualization in a very visceral drag and drop manner. Tableau assimilates with most data types and provides out of the box integration with a variety of big data platforms, comprising Hadoop. Tableau assimilates with R, the business intelligence statistical language many data scientists use to dive deep into BI for statistical and advanced analytics,

Tableau has a ample amount of partner and consultant base, as well as exhaustive online resources including guides, online training and forums. Plus, the Tableau community is well-known enthusiastic and engaging., Like most BI vendors, Tableau provides different software licensing options. There's a nonpaid version of Tableau for personal use, but its scope is limited. To download Tableau to your desktop, you'll pay in the range of \$1000 and \$2000 per year. Online or hosted access to Tableau costs \$500 per person, per year. In the below paragraph the paper has described some cons of tableau server and tool

Hence, lot of organizations will need Tableau server. Based on what companies that have considered Tableau (or have purchased it) tell us, an entry-level server license is about \$1000 per user (with a 10-user minimum). If data is in Excel or a CSV file format, you can upload it to Tableau and perform different data analytical operations Tableau occupied its mark on the world of BI by being one of the first companies to give business customers the ability to operate fairly complex data visualization in a very intuitive, drag and drop manner.. But, if you want to interact to a database, a developer skilled in SQL will have to create the SQL query to retrieve the dataset.

With the Tableau server, there is no term of versioning. You build your reports and dashboards through your desktop and then publish them to the server. Once they're

published, there's no alternate way to retrieve previous versions- once you overwrite there's no pulling back options is available. , In the below figure [1] describes the

excel format of data that is taken for processing in tableau platform.

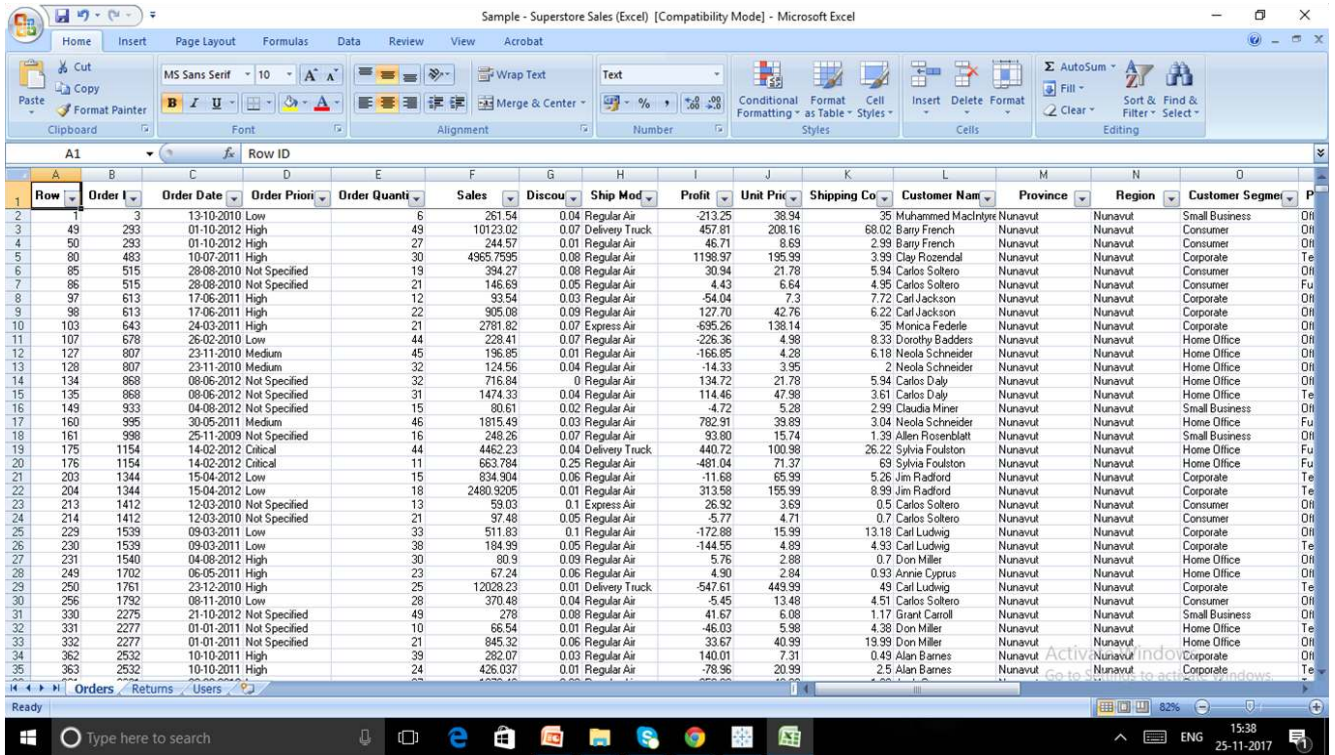


Fig.1. Original Data source in excel sheet

The below figure [2] describes the process of importing the original data source which is present in excel format to tableau-platform.

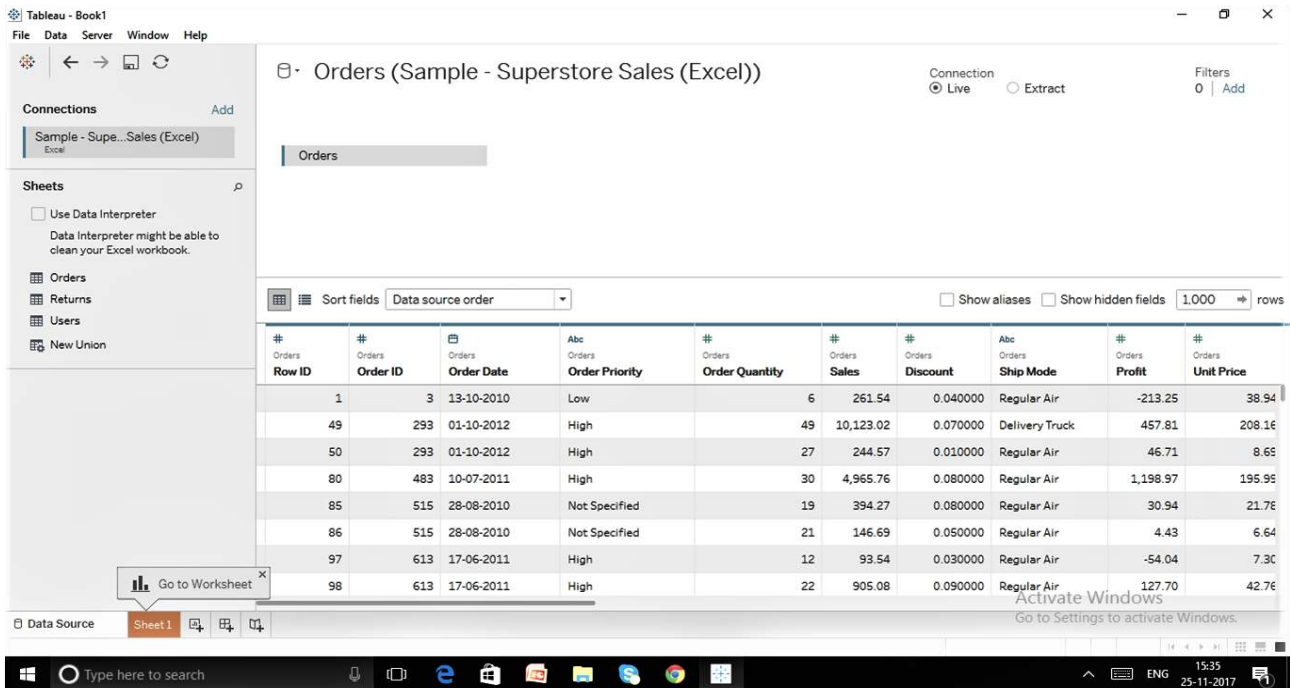


Fig.2. Importing data set in to tableau framework

After successfully importing the data in to tableau platform the data is divided in to two parts the numeric data are generally come in to the measure section and rest other data come to the dimension section. This division of data is generally done by the tableau itself. The a division of data in to dimension and measure section which is done by tableau internally, In the below figure[3] and

Figure [4] the paper describes the data that is present on the original data source ,In the original data set profit and sale information about various, categories of products are present up to year 2017 and the paper has forecasted the profit and sale information on the forthcoming quarters of 2018.

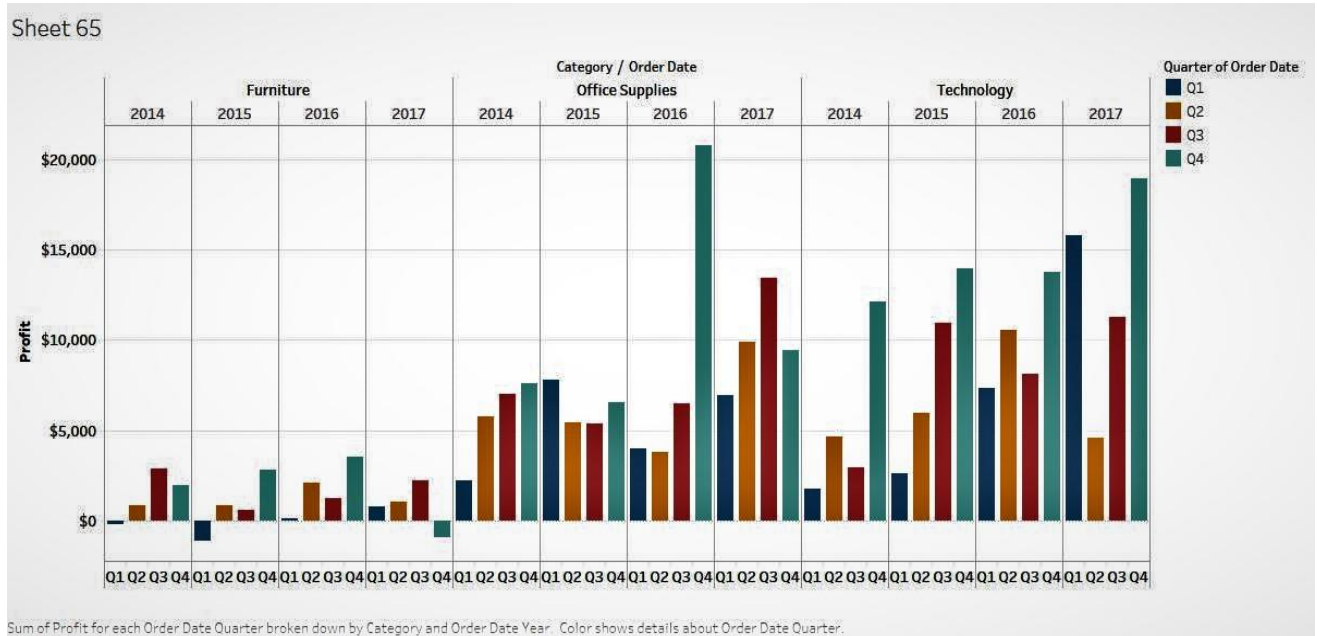


Fig.3. Profit Data Retrieve From Tableau Tool

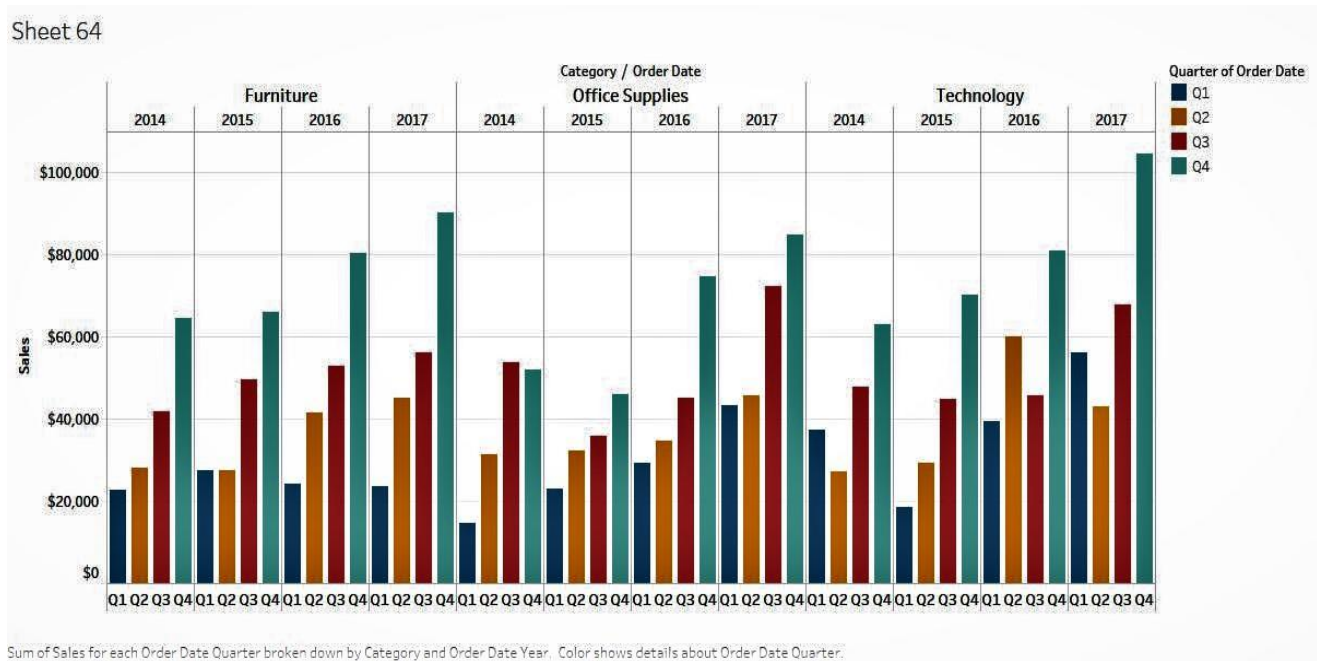


Fig 4.Sales Data retrieve from tableau tool

In the below section the paper has described the estimated profit of different categories of product in forthcoming quarters. In Figure [5] the estimated profit for Furniture Category for forthcoming quarters are calculated. In similar manner Figure [6] provides the estimate calculation of technology category product and Figure [7] represents the estimated calculation Office supplies product category. In tableau the forecast calculation is done on the basis of import data .

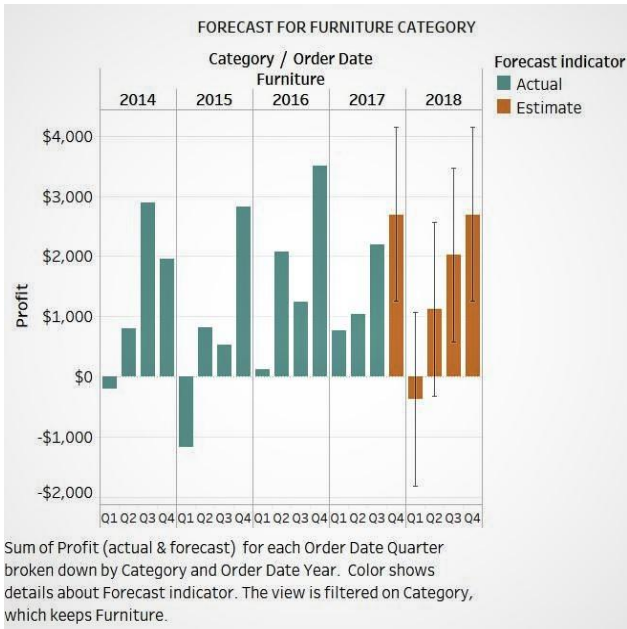


Fig.5. Estimated Profit Calculation For Furniture Category

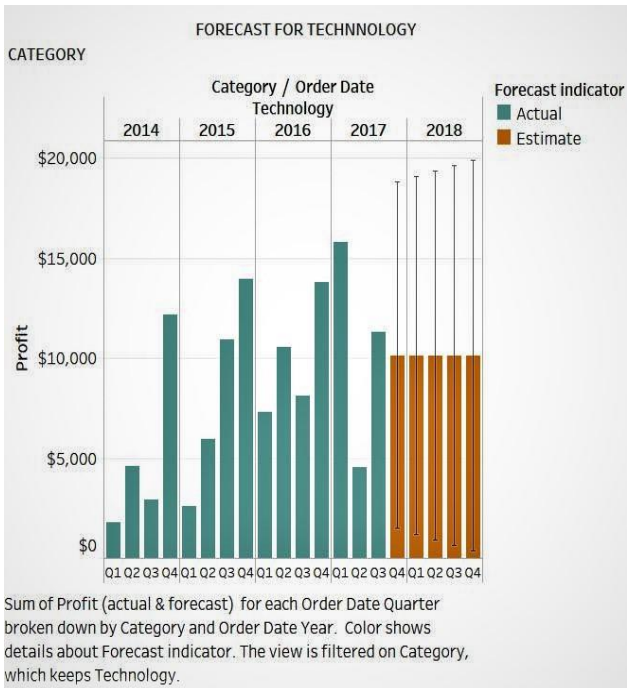


Fig.6. Estimated Profit Calculation For Technology Category

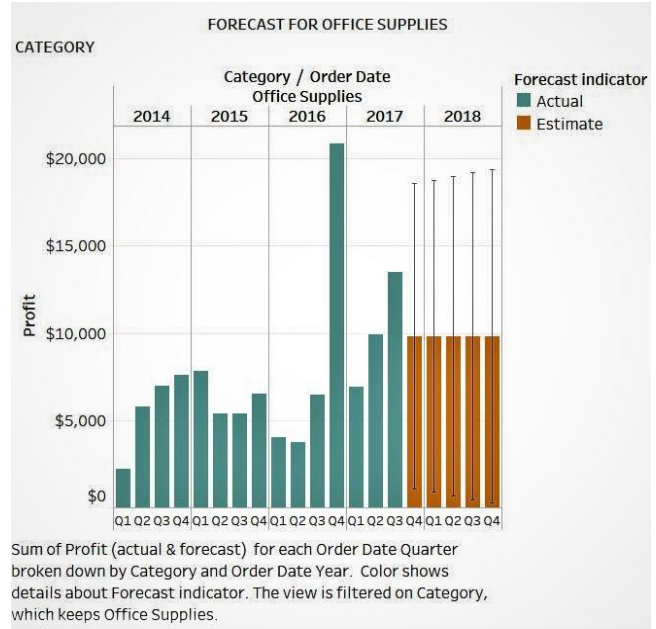


Fig.7. Estimated Profit Calculation For Office Supplies Category

In the below section the paper has described the estimated Sales of different categories of product in forthcoming quarters. In Figure [8] the estimated sale for Furniture Category for forthcoming quarters are calculated. In similar manner Figure [9] provides the estimate sale calculation of technology category product and Figure [10] represents the estimated sale calculation of Office supplies product category, In tableau the forecast calculation is done on the basis of currently available sales data of different quarters. .

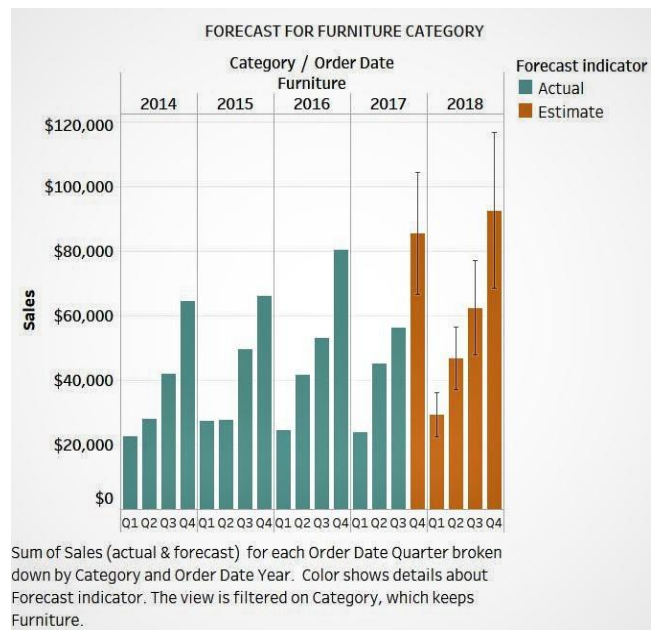


Fig.8. Estimated Sale Calculation For Furniture Category

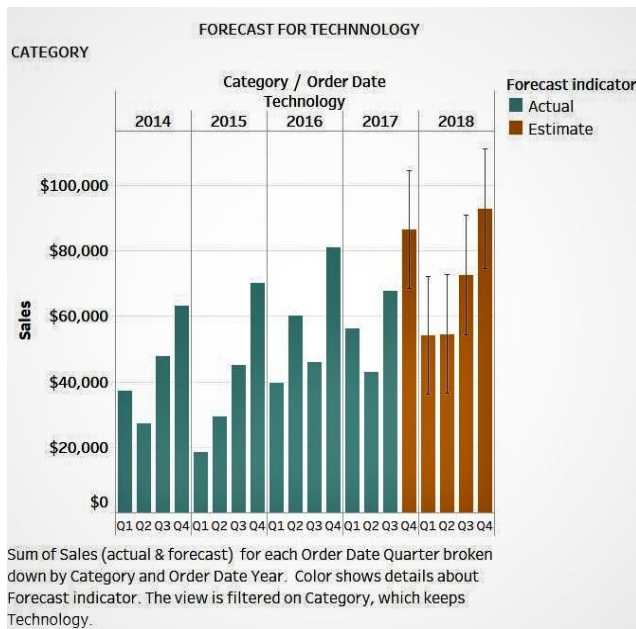


Fig.9. Estimated Sale Calculation For Technology Category

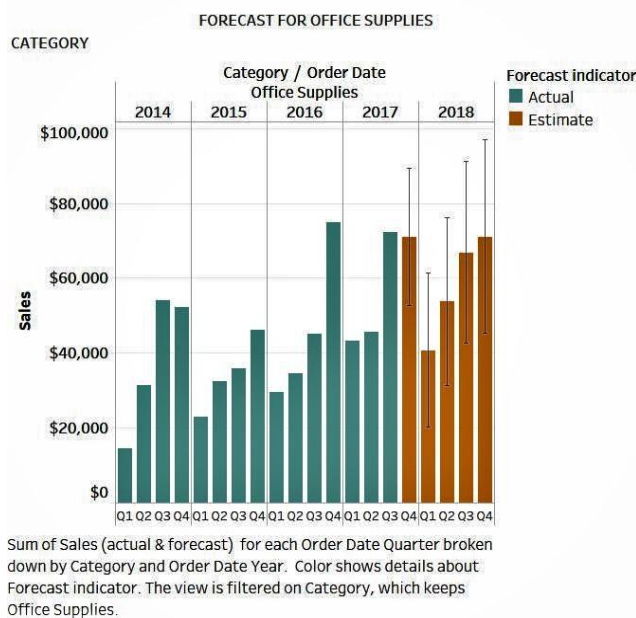


Fig.10. Estimated Sale Calculation For Office Supplies Category

V. CONCLUSION

Due to the increased market competition and increased data management an era has reached that needs more optimization technique to analysis the data. Big data technologies such as HADOOP proposes a framework for processing the data parallel y and after processing the data different visualization tools are used for representing the analyzed result in an effective manner which plays an important for decision making process in different business organizations[11]. Tableau will enhance decision making, add operational awareness, and increase performance throughout the organization[12]. The presented paper describes different tools used for

business intelligence field and provides a depth knowledge regarding the tableau tool. Data visualization plays an important role for the business analytics in order to do different operation with data. Tableau analytics is one of the easiest and most powerful analytics tools today’s era[13]. Tableau can assist to assemble sense of different types of data coming in from multiple sources and blend it into one unique platform so that everybody can make sense of it and utilize it[14].The Tableau Dashboard is an important feature of the Tableau Product that lets you allow to create unique data stories, It is constructed from the ground up for people who do not have any technical skills or coding experience for that matter. So anything can be done by anybody without any prior set of skills. This paper has explained the forecast methodologies in tableau framework and practically done the forecast by taking the dataset from business domain. There are many various types of visualization options available in Tableau which enhance the user experience. Also, Tableau is very easy to learn compared to Python, Business Objects and Domo, anyone without having knowledge of coding can easily learn Tableau. Tableau can operate millions of rows of data with ease. Various types of visualization can be initiated with a huge amount of data without impacting the performance of the dashboards. Also, there is an option available in Tableau where the user can make and can “live” to connections to different data sources like SQL etc and can perform multiple data manipulation operation [16] with the live data. Tableau Dashboard has a marvellous reporting feature that allows you to create personalize dashboard specifically for a convinced device such as a mobile phone or laptop. Tableau automatically figure out which device is the user is viewing the report on and do adjustments to ensure that the right report is delivered to the right device.

Tableau has done a great job climb its way to the top of data visualization tools. So, Tableau has spent more than six years as a leader. However, with the increasing interest in data science, artificial intelligence, and machine learning, Tableau may be left behind if it doesn’t innovate quickly. Tableau does not support the aspect of automatic exhilarating of the reports with the help of scheduling. There is no other choice of scheduling in Tableau. Therefore, there is always some manual effort required when users need to update the data in the back-end. Tableau is not entirely an open tool. Just like other tools like Power BI, developers can create custom visuals that can be easily imported Tableau. So, any new visuals need to be recreated instead of imported. Tableau’s parameters are constant and always unique value can be preferred using a parameter. Whenever the data gets changed, these parameters need to be updated manually every time.

VI. FUTURE SCOPE

Tableau is strictly a visualization tool. Tableau Desktop only permits you to perform very basic preprocessing, Which includes joining and blending

data[17]. Also, you have the competence to change data types. In an ideal world, most data would be exported in perfect tables. However, data cleansing is a necessary step. In most cases, an analyst needs to build a data model with reappearing to format the data. This needs a tool such as Altyrex, Power BI, Python or even Excel to operate data prior to loading[18]. In 2018, introduced their own data preparation tool called Tableau Prep, Different researches are going on for optimizing the execution time and reduce the elapsed time for performing data analytics operation in tableau platform.

REFERENCES

- [1] Raghupathi W: Data Mining in Health Care. In Healthcare Informatics: Improving Efficiency and Productivity. Edited by Kudyba S. Taylor & Francis; 2010:211–223.
- [2] B. Jena, M. K. Gourisaria, S. S. Rautaray, and M. Pandey, "A surveywork on optimization techniques utilizing map reduce framework in hadoop cluster," *International Journal of Intelligent Systems and Applications*, vol. 9, no. 4, p. 61, 2017.
- [3] B. Jena, M. K. Gourisaria, S. S. Rautaray and M. Pandey, "Name node performance enlarging by aggregator based HADOOP framework," *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, 2017, pp. 112-116.
- [4] B. Jena, M. K. Gourisaria, S. S. Rautaray and M. Pandey, "Improvising name node performance by aggregator aided HADOOP framework," *2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, Kumaracoil, 2016, pp. 382-388.
- [5] Nagesh HR, Guru Prasad "High Performance Computation of Big Data: Performance Optimization Approach towards a Parallel Frequent Item Set Mining Algorithm for Transaction Data based on Hadoop MapReduce Framework" *International Journal of Intelligent Systems and Applications(IJISA)*, Vol.9, No.1, pp.75-84, 2017. DOI: 10.5815/ijisa.2017.01.08.
- [6] Siddharth S Rautaray, and Manjusha Pandey, "Single and Multiple Hand Gesture Recognition Systems: A Comparative Analysis", *I.J. Intelligent Systems and Applications*, 6 (11), 57-65, 2014.
- [7] Rath Jairak, Prasong Praneetpolgrang, Nivet Chirawichitchai,"A Roadmap for Establishing Trust Management Strategy in E-Commerce Services Using Quality Based Assessment", *IJIEEB*, vol.6, no.5, pp.1-9, 2014. DOI: 10.5815/ijieeb.2014.05.01.
- [8] Jeffrey Dean and Sanjay Ghemawat , " Map Reduce: Simplified Data Processing on Large Clusters", *IEEE Micro*, 23(2):2228, April 2005.
- [9] Esma Yildirim, Engin Arslan, Jangyoung Kim, Tevfik Kosar. "Application-Level Optimization of Big Data Transfers through Pipelining, Parallelism and Concurrency", *IEEE Transactions on Cloud Computing*, 2016
- [10] J. Archenaa, E.A. Mary Anita," A Survey of Big Data Analytics in Healthcare and Government", *Procedia Computer Science, Elsevier*, Volume 50, 2015, Pages 408–413, Big Data, Cloud .
- [11] Gunjan Varshney1, D. S. Chauhan2, M. P. Dave," Evaluation of Power Quality Issues in Grid Connected PV Systems", *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 6, No. 4, August 2016, pp. 1412~1420.
- [12] Ziv J., Lempel A., "A Universal Algorithm for Sequential Data Compression," *IEEE Transactions on Information Theory*, Vol. 23, No. 3, pp. 337-343.
- [13] N.E. Ayat, M. Cheriet, C.Y. Suen , "Automatic model selection for the optimization of SVM kernels," *Artificial intelligence in medicine*, vol. 38, no.10, pp. 1733-1745, 2005.
- [14] Hamid Bagheri , Abdusalam Abdullah Shaltooiki., " Big Data: Challenges, Opportunities and Cloud Based Solutions", *International Journal of Electrical and Computer Engineering (IJECE)*,2015.
- [15] Sabyasachi Chakraborty, Kashyap Barua, Manjusha Pandey, Siddharth Rautaray," A Proposal for High Availability of HDFS Architecture based on Threshold Limit and Saturation Limit of the Namenode", *International Journal of Information Engineering and Electronic Business(IJIEEB)*, Vol.9, No.6, pp. 27-34, 2017. DOI: 10.5815/ijieeb.2017.06.04.
- [16] Mayank Bhushan , Monica Singh , Sumit K Yadav , " Big Data query optimization by using Locality Sensitive Bloom Filter ", *IJCT*, 2015.
- [17] Liu, Yunxiang, and Jiongjun Du. "Parameter Optimization of the SVM for Big Data", 2015 8th International Symposium on Computational Intelligence and Design (ISCID), 2015.
- [18] Lanchao Liu and Zhu Han , " Multi-Block ADMM for Big Data Optimization in Smart Grid " , IEEE, 2015.
- [19] Al-Madi, Nailah, Ibrahim Aljarah, and Simone A. Ludwig. "Parallel glowworm swarm optimization clustering algorithm based on MapReduce", 2014 IEEE Symposium on Swarm Intelligence, 2014.
- [20] A. Ramaprasath, K. Hariharan, A. Srinivasan, "Cache Coherency Algorithm to Optimize Bandwidth in Mobile Networks", Springer Verlag, Lecture Notes in Electrical Engineering, Networks and Communications, Chapter 24, Volume 284, 2014, pp 297-305.
- [21] Ziv J., Lempel A., "A Universal Algorithm for Sequential Data Compression," *IEEE Transactions on Information Theory*, Vol. 23, No. 3, pp. 337-343
- [22] E. Yildirim, J. Kim, and T. Kosar, "Optimizing the sample size for a cloud-hosted data scheduling service," in Proc. 2nd Int. Workshop Cloud Computing. Sci. Appl., 2012.
- [23] 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
- [24] Abdus Satter, Nabil Ibtehaz," A Regression based Sensor Data Prediction Technique to Analyze Data Trustworthiness in Cyber-Physical System", *International Journal of Information Engineering and Electronic Business(IJIEEB)*, Vol.10, No.3, pp. 15-22, 2018. DOI: 10.5815/ijieeb.2018.03.03

Author's Profile



Bibhudutta Jena, is a CSI Accredited Student. Completed M. Tech (Computer Science and Engineering) at the School of Computer Engineering, KIIT University, Bhubaneswar. His areas of interest Data Analytics ,Data mining,Data visualization etc .He has already published more than 10 research papers in big data and analytics domain,He is currently pursuing as an software

engineer in MPHASIS Limited. He can be reached at bibhudutta93@gmail.com.

How to cite this paper: Bibhudutta Jena, "An Approach for Forecast Prediction in Data Analytics Field by Tableau Software", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.11, No.1, pp. 19-26, 2019. DOI: 10.5815/ijieeb.2019.01.03