

FORECAST OF SALES OF WALMART STORE USING BIG DATA APPLICATIONS

Anita S. Harsoor¹, Anushree Patil²

¹Assistant Professor, Dept. of Computer Science and Engineering, Poojya Doddappa appa College of Engineering, Karnataka, India

²M. Tech, Dept. of Computer Science and Engineering, Poojya Doddappa appa College of Engineering, Karnataka, India

Abstract

Many household products are sold by various subsidiaries of the retail store network which are geographically located at various locations. Supply chain inefficiencies will occur at different locations when the market potential will not be evaluated by the retailers. Many times it is not easy for the retailers to understand the market condition at various geographical locations. The organization of retail store network has to understand the market conditions to intensify its goods to be bought and sold so that many number of customers get attracted in that direction. Business forecast helps retailers to visualize the big picture by forecasting the sales we get a general idea of coming years if any changes are needed then those changes are done in the retail store's objective so that success is achieved more profitably. It also helps the customers to be happy by providing the products desired by them in desired time, when the customers are happy then they prefer the store that provides all the resources they need to their satisfaction by this the sales in the particular store in which the customers purchase more items increases causing more profit. The forecasting of sales helps to know the retailers the demand of the product. In this paper we make an attempt by understanding the retail store business's driving factors by analyzing the sales data of Walmart store that is geographically located at various locations and the forecast of sales for coming 39 weeks is done. By sales forecasting the retail networks are supported so that the resources can be managed efficiently.

Keywords: Hadoop, Hive, Big data, Map reduce, Holt winters.

1. INTRODUCTION

We all are very curious about our future! very excited to know what will happen with all of us the very next moment, tomorrow, similarly the retailers also curious about their business, its progress, curtailing factors of it. By adopting certain steps the aspects that can cause damage or reduce the profit can be avoided. In this goal of forecasting the sales business the data from the various sectors is collected and the data analytics is done the efficient understanding of the observed data by common steps is not practically possible because the data is very huge the masses of data of organization is molded in such a way that its having meaning by understanding deeply the suitable actions can be taken. Let us consider the retail stores network example Walmart it is an example for giant stores, big bazaars etc.

The retail store sells the household products and obtains profit by that. There are different subsidiaries of the retail store network whose locations are variously located at various geographical locations most of the time retailers will not be successful in understanding the customer's needs because they will be able in the evaluation of market potential at that location, during special occasions the rate of sales or shopping is more sometimes this may cause inefficiency of the products, the relationship between the customers and the stores is analyzed and the changes that need to obtain more profit is done. The history of purchase of each product in each store and department is maintained

by observing these sales are forecasted which enables the knowledge of profit and loss occurred during that year. In particular department during the particular session let us consider example Christmas. During Christmas festival the sales is more in department like clothing, foot wears Ref[6] etc., then during summer the sales of cotton clothes is more, during winter the sales for Sweaters is more. The sales of products changes as per the session by observing this history of sales, the sales can be predicted for the future. That finds the solution of uneasiness in the business of retail store network. Supply chain management[4] is the advantage of competition, the main majors of supply change management are to increase the profit of sales and to manage the inventory turnovers, when the supply change are observed properly then a clear picture is obtain about a particular store whether there is a profit from that store are its under loss. Then accordingly suitable operations are done to be successful. Here the retailers observe the customers and they target them by some attractive offers. So that they will be back to the store and spend long time and more money. In this paper we forecast the sales by using three modules they are hive, R programming and tableau, storage in hive is large hive component is Horton/network data platform. SQL is provided by hive that provides interface to the data stored in HDP hive used for data processing and understanding in hive partitioning and bucketing is done. In R programming statistical is done in tableau the interactive visualization of the data on the

product that are focused on the business. The extraordinary business advantage and the good service delivery are the gifts of big data. The positive changes in the different business organization's decision are made by big data [9]. The term big meaning it's a data set that arise the difficulties in managing the data sets by the use of existing concepts of management and tools. The most significant role played in big data is map reduce. The important properties of map reduce are its elasticity, scalable when the analysis of large data sets is done the map reduce can tolerate the fault and is efficient. The features of map reduce when compared with other design models are more flexible which makes map reduce more popular in the large scale data processing.

2. BACKGROUNDS

2.1 Fundamentals

The main aim of retail store network planning is to obtain maximum profit by knowing the information and where to invest profitably and in which stores. To be successful the organizations should be able to know the customers of each store in an assigned region, the several combinations of the location of allocation options, the options are evaluated with certain models of mathematics or algorithms this is helpful in final decision of the configuration of store network with the perfect strategy they are the format of store, capacity of the store and goods to be bought and sold etc. The different types of store format could be convenient store, department store, luxury store, super market, shopping malls etc. helps in determining the business models and strategy of operations [1].

3. CHALLENGES AND MOTIVATIONS

There are many challenges in the retail store network planning some of them are retailers fail in the evaluation of the potential of the market. Retailers ignore the seasonal randomness. The supply chain inefficiencies when the products have great demand then they are not available. The human resources are inefficient the employees are not available whenever necessary. The retailers face the difficulties in inventory management system; sometimes the retailers ignore the competition in the market.

Retailers develop the plans that promotes the success and the highly target plan. The plans should be such that they help to obtain the maximum profit. The new product lines should be developed or they should be purchased with confidence. The supply chain mechanism should be efficient.

3.1 Forecasting of Market Potential

The overall turnover of the market of a particular item are store which is the result of estimating the highest possibilities of volume of sales in given period of time and under certain assigned conditions. Market potential depends on the customer's behavior, the opponents, facilities support etc. These influence the sales of future of a particular store of particular location.

3.2 Planning of Store

Store planning is important and is systematic strategy process in both firm level and individual store level. Goods to be bought and sold, supply chain management and human resource management, space management main work in planning of store. By observing the history of the stores it helps to get an idea of sales of the store and make any changes in the objective so that it can be more profitable. The basic information provided by the existing store is very important in the prediction of sales. In this paper we considered the data of walmart store. The key data sets that should be used to analyze and report on in the retail industry are:

Sales Data - Point of sales data, gross margins, turnovers, gross margin return on inventory investment.

Market Data - Market share, competitors pricing, competitors product lines, competitors market share and customer profit.

Promotional and Marketing Data - Success of past, customer's who provide the promotion of sales their feedback is observed, total cost of promotion. Pricing offers

4. IMPLEMENTATION OF SALES FORECASTING

4.1 Sales Forecasting System Architecture

Architecture is the most important aspect of any system. Fig.1 represents the system architecture.

4.2 The Forecasting Process

The process of forecasting [15] is a group of methods to predict the sales. It is initiated after determining the objective. It may include the sales amount in dollars, the number of employees to be appointed. In Fig.2 selection of dependent and independent variables are done. The forecasting results like sales data or the number of employees to be appointed in the upcoming year. The market factor include the factors like products existence in the store, its quality and the demand of the item. Market index is a market factor that is expressed as the amount of percentage relatively with some base content. When the market index is increased then the industry sales is increased. The index consists of many market factors like price, population of the area, personal income that is disposable. Then in the forecasting process the procedures of forecast and methods that are useful for data analysis are determined. If the procedures were not used in prior then the firm may want to test the procedures. Then gathering and analyzing of data is done. Certain assumptions are made about the forecasted sales. Then the sales forecast is finalized as the time passes and the results are evaluated.

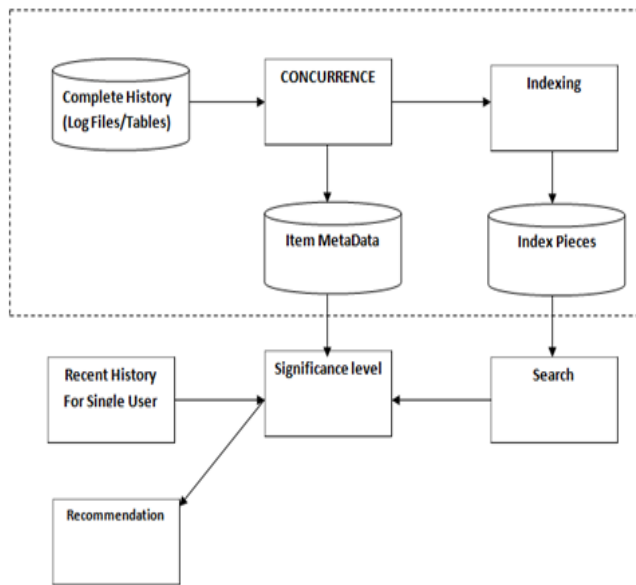


Fig-1: System architecture

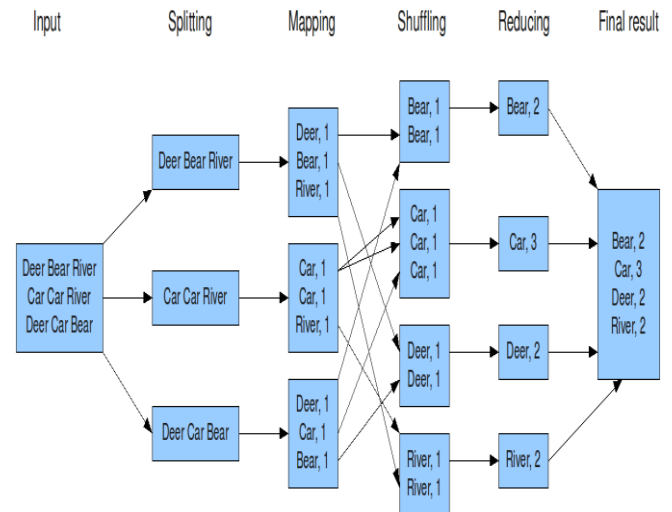


Fig-3: Representation of MapReduce

consisting of the repeated words that general sentence is split into number of words then the code or numerical representation for each word is assigned this is called as mapping after this the words are shuffled so that all the repeated words are together then reducing is done to obtain the final result. Fig.4 represents the Big data, buzz word Big data is used in description of huge data volume including structured and unstructured by using traditional database processing is difficult. Big data is used in the business, the accurate results obtained by Big data helps to make decision about the results, in the prediction of analytics Big data is helpful. Big data is used by many people like scientists, business people, government. Fig.5 represents the overview of Hadoop. Apache software foundation released open source software called as Hadoop which operates on commodity hardware, it is processing framework, distributed storage and parallelized. Hadoop has various server nodes that are useful in the storage and parallel processing. In massive data sets the execution of batch processing is allowed by hadoop programming model Map reduce is used to generate vast data sets intermediate results/key values are obtained in map reduce all these intermediate values are merged by reduce function by the intermediate key to obtain the results.

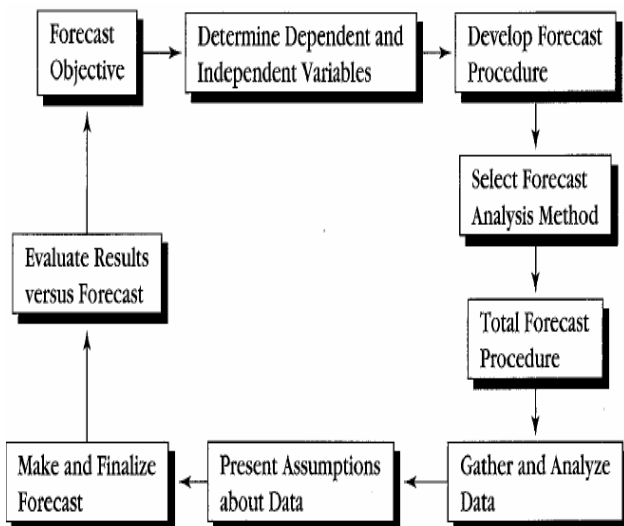


Fig-2: The forecasting process

5. METHODOLOGY

Map reduce: To generate huge amount of data sets with parallel, distributed algorithm on a cluster programming model called Map reduced[5] is used for processing. In map reduce [14] data is focused rather than algorithms. The large data is necessary for processing. In this paper we consider the sales data of walmart store for three years having 45 stores and each store has 99 departments in the various locations. Observation of walmart store’s sales data is done for three years and the sales forecast is done for the next 39 weeks. In fig.3 we can see the Map reduce diagram, we consider a general sentence



Fig-4: Representation of Big data

5.1 Strategic Research

The strategy includes the collection of huge data of sales and then it is transferred on HDFS (Hadoop distributed file system)[10] and map reduced is performed on the data sets. The data after processed by map reduced though is in understandable format, but is too large hence it is difficult to draw conclusions out of that. Processing using hive [11] is done by loading the data sets supplied then dynamic partition and bucketing is applied. The monthly sales of 45 stores and 99 departments are calculated. The average feature across them is calculated.

5.2 Machine Learning

Machine learning algorithm which provides efficient results are applied so that huge data can be modeled R is a programming language and software environment for statistic computing and graphs. The R programming [12] is used by the statisticians and data miners widely to develop the statistical software and data analysis. The Holt winters [16] algorithm is used to predict the sales. The seasonality, trend and randomness is observed in the algorithm. The algorithm is used for train data sets and then the sales prediction is done.

5.3 Graphical Representation

The processed data by the machine learning algorithm which provides efficient results can be understood but this obtained information is too large and hence conclusions are difficult. We all know the picture representation is worth than many words. Hence this information can be understood by the tableau's interactive data visualization focused on business intelligence. Tableau [13] helps in interactive data visualization of products.

6. RESULTS AND DISCUSSION

Results and discussions are very important so that readers can understand the information of the paper, results are the descriptive content information. Below are the few results obtained by observing these results the sales is forecasted for the next year. Fig-6 represents the different combinations of the graphs. Fig-6 a represents the sales and stores, x axis represents the stores and y axis represents the weekly sales. Fig-6 b the store, department and sales it's a bubble chart, Fig-6 c weekly sales and the department, x axis represents the weekly sales and y axis represents the departments, suppose if we want the sales of department in store 1 then we can select the required department and the sales is represented. Fig-6 d represents the weekly sales and the year, the weekly sales according to year is represented, x axis represents the year and y axis represents the weekly sales. Fig-7 represents the sales and the stores x axis represents the different number of the stores and the y axis represents the sales in each department. Fig-8 represents the sales in stores by using Tableau, it is the bubble chart. Fig-6, Fig-7 and Fig-8 are the graphs of the data collected from 3 years of walmart store, now we have to predict the sales for the next 39 weeks and should be graphically represented. This is done by using the Holt winters algorithm and the forecasted sales is represented in Fig.9 x axis represents the years and y axis represents the sales, by observing the Fig.9 we will not get the clear picture of the forecasted sales hence to get a clear knowledge of the forecasted sales we represent it numerically, Fig-10 represents the numerical representation of the forecasted sales and the accuracy of sales predicted is measured by 80% low confidence sales, 80% high confidence sales and 95% low confidence sales and 95% high confidence sales, error factor can be found between the predicted sales and the observed sales data, let us consider with an example, if we want to find the error factor of month June in both the predicted sales and the observed sales data then we can know it by difference between predicted sales and the observed sales data, if the difference between them is very low or negligible then the sales predicted is accurate.

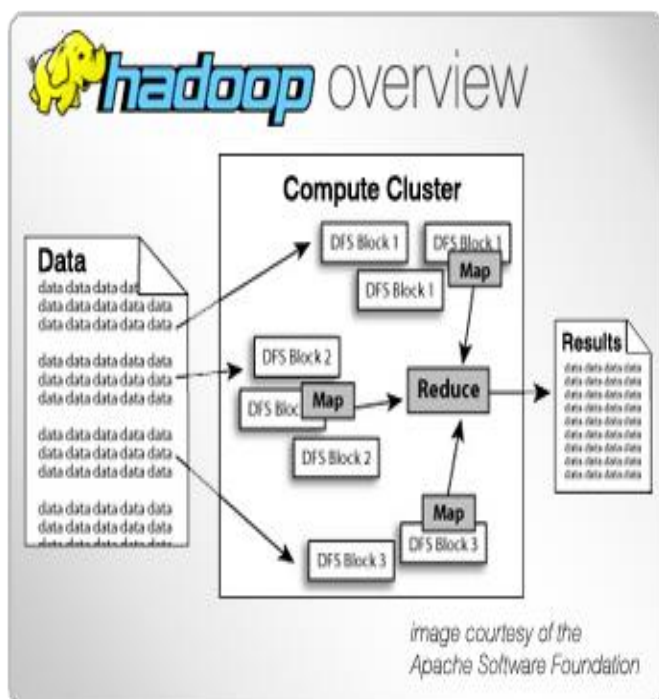


Fig-5: Representation of overview of Hadoop

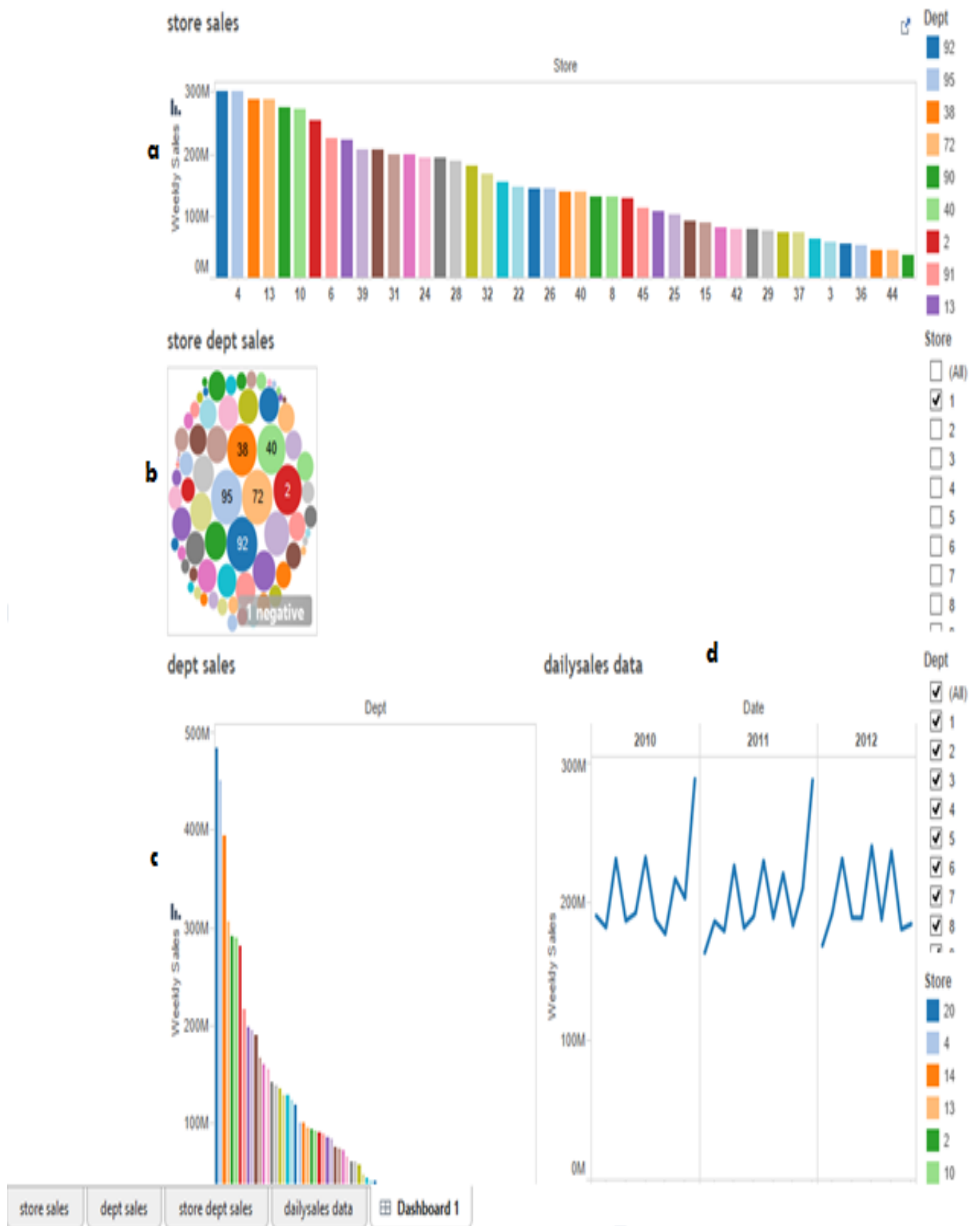


Fig-6: the combination of all graphs, Fig-6: a represents sales and store, Fig-6: b represents store, departments, sales bubble chart, Fig-6: c represents weekly sales and departments, Fig-6: d represents weekly sales and year (daily sales)

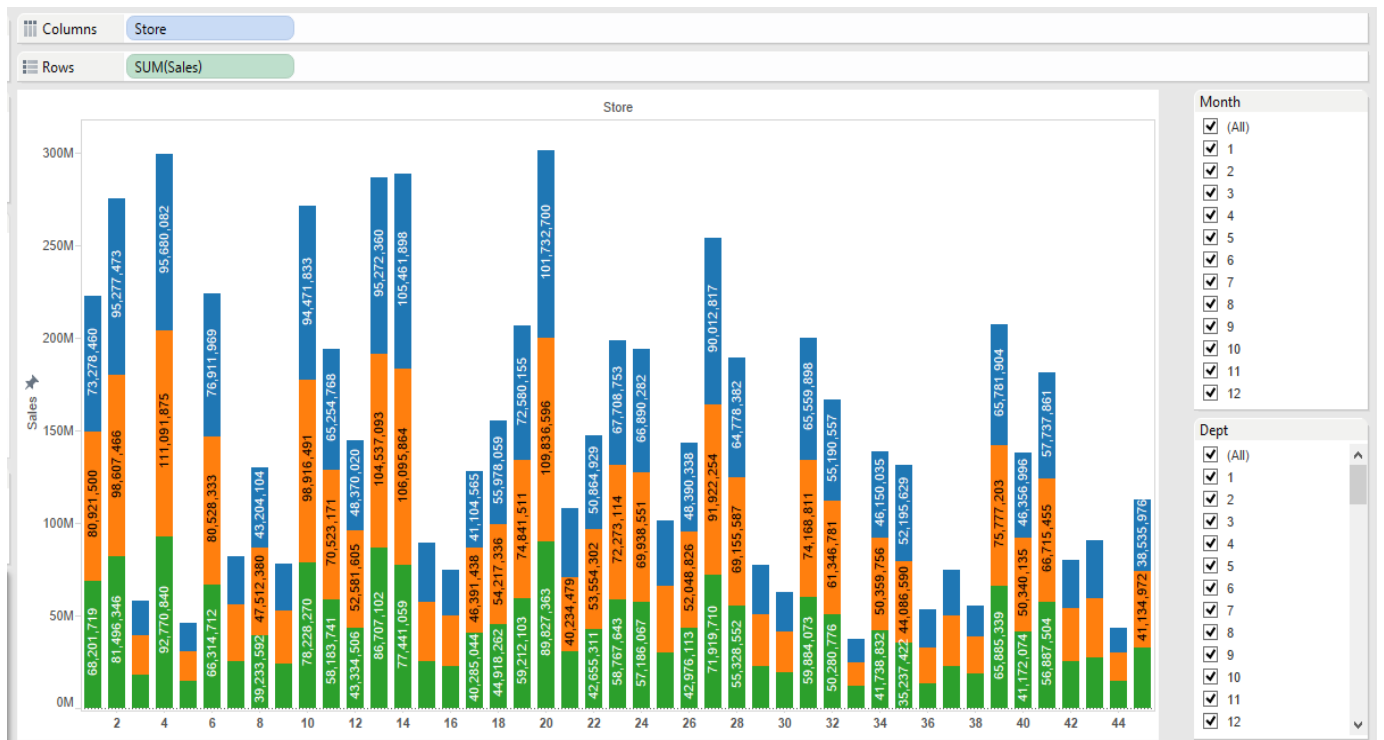


Fig-7: the sales and stores in Tableau

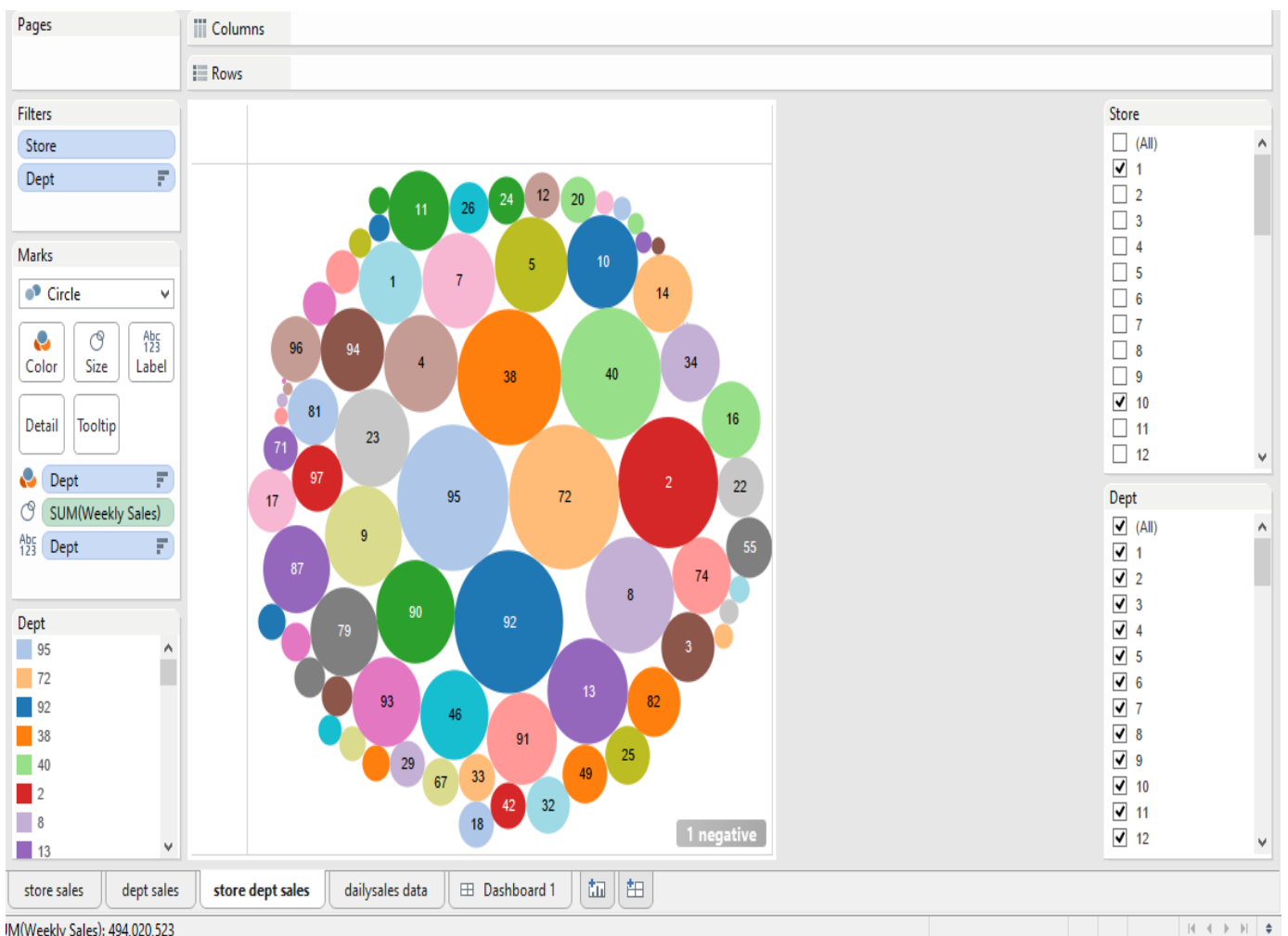


Fig-8: the representation of sales in stores by bubble chart in Tableau

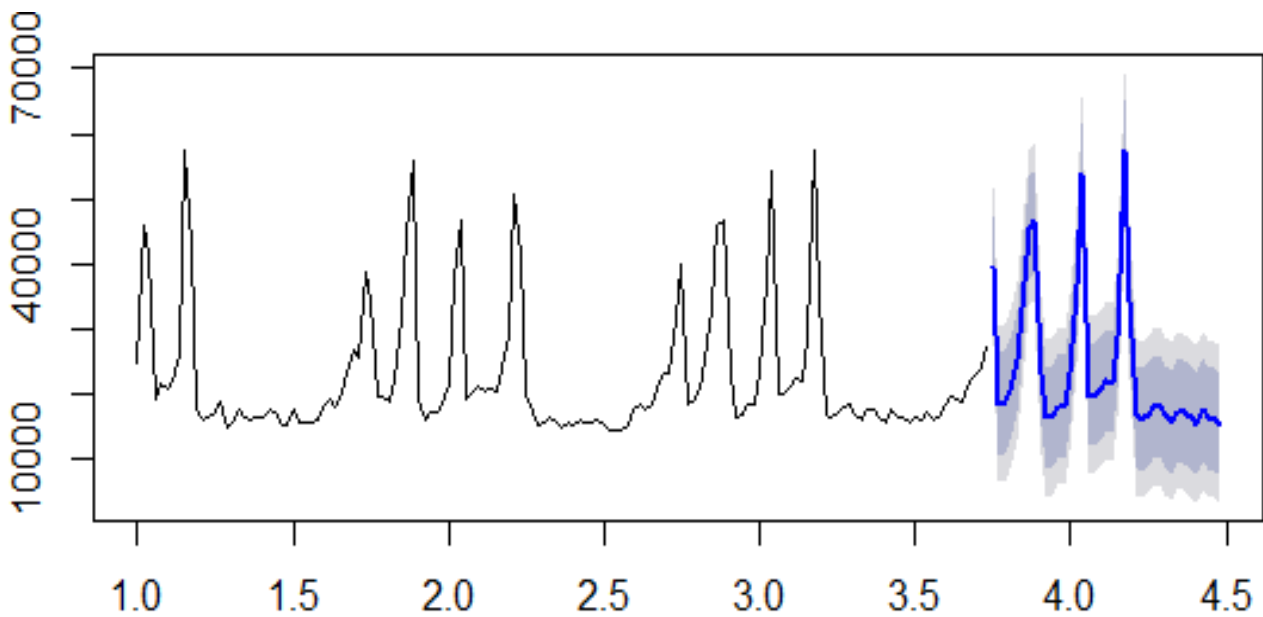


Fig-9: the predicted graph obtained by Holtwinters

Console Output:

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
3.750000	39641.87	31756.06	47527.69	27581.562	51702.19
3.769231	18445.35	10559.54	26331.17	6385.042	30505.67
3.788462	18806.47	10920.66	26692.29	6746.162	30866.79
3.807692	20667.06	12781.25	28552.88	8606.752	32727.38
3.826923	25049.30	17163.49	32935.12	12988.992	37109.62
3.846154	33061.73	25175.92	40947.55	21001.422	45122.05
3.865385	45528.84	37643.03	53414.66	33468.532	57589.16
3.884615	46544.56	38658.75	54430.38	34484.252	58604.88
3.903846	23106.69	15220.88	30992.51	11046.382	35167.01
3.923077	16323.50	8437.69	24209.32	4263.192	28383.82
3.942308	16650.21	8764.40	24536.03	4589.902	28710.53
3.961538	18120.91	10235.10	26006.73	6060.602	30181.23
3.980769	18133.97	10248.16	26019.79	6073.662	30194.29
4.000000	23266.30	15380.49	31152.12	11205.992	35326.62
4.019231	36744.30	28858.49	44630.12	24683.992	48804.62
4.038462	53815.91	45930.10	61701.73	41755.602	65876.23
4.057692	19880.03	11994.22	27765.85	7819.722	31940.35
4.076923	19868.84	11983.03	27754.66	7808.532	31929.16
4.096154	20895.88	13010.07	28781.70	8835.572	32956.20
4.115385	22122.69	14236.88	30008.51	10062.382	34183.01
4.134615	21863.51	13977.70	29749.33	9803.202	33923.83
4.153846	28708.67	20822.86	36594.49	16648.362	40768.99
4.173077	57347.93	49462.12	65233.75	45287.622	69408.25
4.192308	34440.02	26554.21	42325.84	22379.712	46500.34
4.211538	16732.00	8846.19	24617.82	4671.692	28792.32
4.230769	16103.41	8217.60	23989.23	4043.102	28163.73
4.250000	16903.25	9017.44	24789.07	4842.942	28963.57
4.269231	17920.01	10034.20	25805.83	5859.702	29980.33
4.288462	18273.60	10387.79	26159.42	6213.292	30333.92
4.307692	16719.36	8833.55	24605.18	4659.052	28779.68
4.326923	15821.30	7935.49	23707.12	3760.992	27881.62
4.346154	17421.81	9536.00	25307.63	5361.502	29482.13
4.365385	17314.63	9428.82	25200.45	5254.322	29374.95
4.384615	16389.22	8503.41	24275.04	4328.912	28449.54
4.403846	15478.63	7592.82	23364.45	3418.322	27538.95
4.423077	17579.18	9693.37	25465.00	5518.872	29639.50
4.442308	16321.99	8436.18	24207.81	4261.682	28382.31
4.461538	16103.87	8218.06	23989.69	4043.562	28164.19

Environment:

- decom: List of 6
- h1: List of 9
- i: 1L
- j: 4L
- model: List of 9
- num: int [1:143] 6 7 8 9 10 11 12 13 14 15 ...
- num1: int [1:39] 45 46 47 48 49 50 51 52 1 2 ...
- pred: List of 10

Fig-10: the predicted numerical output by Holtwinters

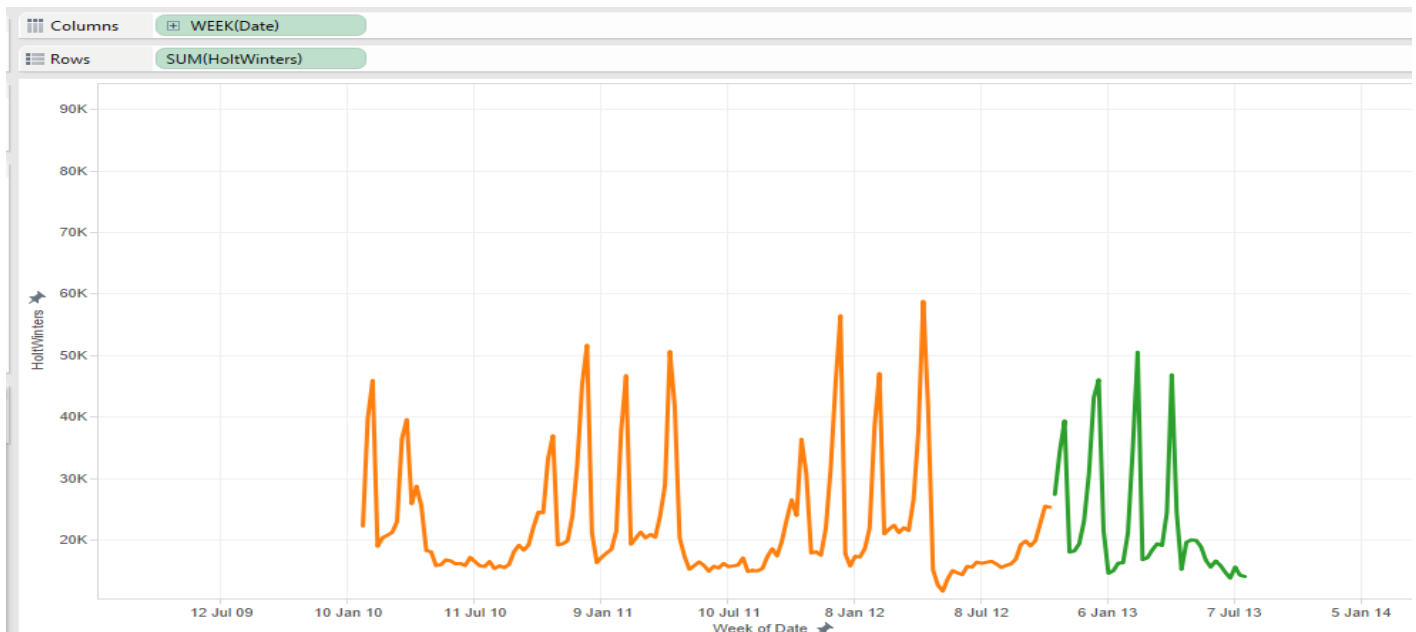


Fig-11: The sales prediction by Holtwinters using Tableau

7. CONCLUSION

Most of the shopping malls / shopping centers plan to attract the customers to the store and make profit to the maximum extent by them. Once the customers enter the stores they are attracted then definitely they shop more by the special offers and obtain the desired items which are available in the favorable cost and satisfy them. If the products as per the needs of the customers then it can make maximum profit the retailers can also make the changes in the operations, objectives of the store that cause loss and efficient methods can be applied to gain more profit by observing the history of data the existing stores a clear idea of sales can be known like seasonality trend and randomness. The advantage of forecasting is to know the number of employees should be appointed to meet the production level. Sales drop is bad thing forecasting sales helps to analyze it and it can overcome through the sales drop to remain in the competition forecast plays a vital role.

REFERENCES

- [1]. A Service-Oriented Solution for Retail Store Network Planning. Xinxin Bai, Wei Shang, Wenjun Yin, Jin Dong.
- [2]. A sales forecasting model for consumer products based on the influence of online word-of-Mouth. Ching-Chin Chern, Chih-Ping Wei, Fang-Yi Shen, Yu-Neng Fan.
- [3]. Using clustering to improve sales forecasts in retail merchandising. Mahesh Kumar · Nitin R. Patel.
- [4]. Simulation Based Sales Forecasting On Retail Small Stores. Hai Rong Lv Xin Xin Bai Wen Jun Yin Jin Dong.
- [5]. Map Reduce for Data Intensive Scientific Analyses. Jaliya Ekanayake, Shrideep Pallickara, and Geoffrey Fox.
- [6]. Prediction of retail sales of footwear using feed forward and recurrent neural networks. Prasun Das & Subhasis Chaudhury.
- [7]. <http://www.yellowfinbi.com/YFCCommunityNews-Data-Analysis-for-the-Retail-Industry-Part-1-100068>

- [8]. <http://www.milanor.net/blog/?p=853>
- [9]. http://www.slideshare.net/Hadoop_Summit/predictive-analytics-with-hadoop
- [10]. http://en.wikipedia.org/wiki/Apache_Hive
- [11]. [http://en.wikipedia.org/wiki/Hue_\(Hadoop\)](http://en.wikipedia.org/wiki/Hue_(Hadoop))
- [12]. [http://en.wikipedia.org/wiki/R\(programming_language\)](http://en.wikipedia.org/wiki/R(programming_language))
- [13]. http://en.wikipedia.org/wiki/Tableau_Software
- [14]. <http://en.wikipedia.org/wiki/MapReduce>
- [15]. Forecasting market demand and sales
- [16]. Holt-Winters Statistical Forecasting and ACO Metaheuristic for Traffic Characterization
- [17]. www.google.com
- [18]. P. A. Morrison and A. F. Abrahamse, "Applying Demographic Analysis to Store Site Selection", Population Research and Policy Review, 15, the Netherlands, 1996, pp. 479-489.
- [19]. ReVelle C.S. and Eiselt H.A., "Location Analysis: A Synthesis and Survey - Invited review", European Journal of Operational Research, Vol.165, No.1 2005, pp.1-19.
- [20]. Jaramillo, J.H., Bhadury, J. and Batta, R., "On the Use of Genetic Algorithms to Solve Location Problems" Computers & Operations Research, 29(6), 2002, pp. 761-79.
- [21]. Aurenhammer F., "Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure", Computing Surveys, 23 (3), SEP 1991, pp: 345-405.
- [22]. O'Kelly M.E., "Trade-area Models and Choice-based Samples: Methods", Environment and Planning, A 31 (4), APR 1999, pp. 613-627.
- [23]. Service-Oriented Enterprise: The Technology Path to Business Transformation.", <http://www.intel.com/business/bss/technologies/soe/> accessed at Jan. 15th, 2007.
- [24]. F. Leymann; D. Roller; M.-T. Schmidt, "Web Services and Business Process Management", IBM Systems Journal, VOL 41, NO 2, 2002, pp. 198-211.

BIOGRAPHIES

Anita S. Harsoor, Asst. Prof. in Department of Computer Science and Engineering, P.D.A. College of Engineering, Gulbarga, Karnataka, India. She has obtained M.Tech (I.T) degree in 2010 and pursuing Ph.D. from Gulbarga University, Gulbarga. Her research interest is Digital Image Processing. She has published research papers in reputed journals and presented papers in conferences.



Miss. Anushree Patil has completed B.E. (CSE) from PDA College of Engineering and now pursuing M.Tech (CSE) from PDA College of Engineering, Gulbarga, Karnataka. Her academic interest area is networking, image processing and big data applications.