ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.7, 5049–5056
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.7.70 Spec. Iss. on AMABDA-2020

AN APPROACH OF SALES PREDICTION SYSTEM OF CUSTOMERS USING DATA ANALYTICS TECHNIQUES

K. VENGATESAN[1], E. VISUVANATHAN, A. KUMAR, S. YUVARAJ, AND P. SHIVKUMAR TANESH

ABSTRACT. In country there are lots of supermarkets, shopping malls the managing sales of the shopping malls is very exigent works. The customer requirements will be changed every month, the manager or store owners are facing the problem for keeping stock for upcoming months. In this work we proposed data analytics-based approach for predicting the sales details based on the last 12 months data and it will generate report, in which moth maximum products are sales profit of the each and every month. This system performs various task like, finding best months for sales, calculate monthly earned money from different products, which city sold the maximum products, when will be the best time for advertisement to maximum customer to buy the products also generates report, which are items are frequently bought by the customer to keep stock for upcoming months and, increase the profit of the shopping mall business.

1. INTRODUCTION

Sales prediction is playing creating and noteworthy employment in various fields, for instance, money related envisioning, electric power deciding, resource conjecture, etc. Arrangements conjecture is a noteworthy basic for enormous business orchestrating and right powerful, allowing associations to all the more

[1]*corresponding author*

2010 *Mathematics Subject Classification.* 60G25, 62M20.

Key words and phrases. Sales Prediction, Maximum Profit, Customer Behaviours, Data Analytics.

promptly structure their business works out. Sales prediction is critical for separated associations, especially vehicle arrangements, land, and other customary undertakings. The desires are generally done by applying verifiable systems, for instance, backslide or the autoregressive–moving-typical (ARMA) taking into account recorded arrangements data. Regardless, these procedures simply work for explicit data. Such enormous quantities of components with complex interrelationships sway bargains and in all likelihood consolidate ones with a sensible degree of defencelessness. Using data mining, we can perceive potential models and headway consistency from most of the data. As needs are, a growing number of researchers revolve around how to use data mining to process recorded data and handle inclines in bargains desire.

As of now, bargains figure investigation on online arrangements has been less analysed because of the lack of certified data with respect to the issue. With the reputation of sagacious compact terminals, E-exchange, especially B2C (Business-to-Customer), has been impacting starting late. Thusly, the fitting arrangements desire methodology in the field of E-exchange to propel profitability in online arrangements errands is a basic issue. In assessment with detached arrangements, online business has its own business qualities, for instance, separated basic customer information and Web program information. The paper proceeds with another perspective that revolves around how to pick a reasonable method to manage guess manage higher reasonability and continuously accurate precision. The data for this paper has been given by a remarkable, genuine Chinese electronic shopping association that is a bit of the B2C exhibit in online business book bargains. We jump into another investigation field, electronic business, and apply veritable arrangements data to a couple of conventional desire models, wanting to locate a trigger model that could pick the correct assessing model to foresee arrangements of a given thing. There is no vulnerability that it will feasibly support an undertaking in choosing bargains decisions in real exercises. It will propel the speculative reason and research techniques far out of tremendous data.

References on the topic are given in [1-10].

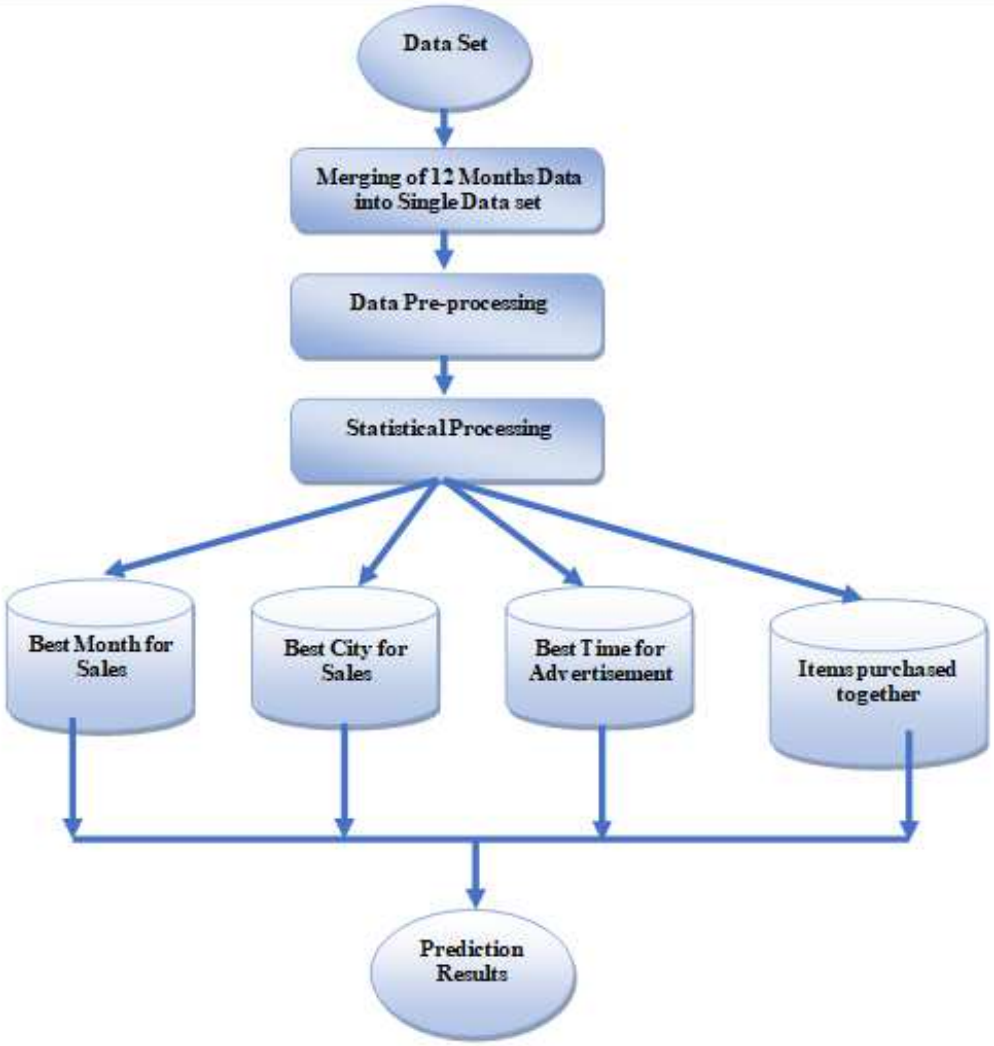

FIGURE 1. Workflow Diagram

2. WORKFLOW DIAGRAM

The Fig.1 demonstrate the work flow model of the proposed sales prediction system, initial stage which collect 12 months data set in to individual files, then which are combined into single data set. Second phase is data pre-processing, which remove the noisy, error values, irrelevant values from the values from the sales dataset. Third phase is statistical processing, which will calculate mean, median, mode, standard deviation of every attributes, and identify how one attribute is related to another attributes. The fourth phase is need to generate

different results from the data set like, best month for sales, best city for sales, best time for advertisement, list of items is purchase by customer together and finally need to find the reason for particular item purchased by many customers. Our source data set contains following attributes like Order ID, Product, and Quantity Ordered, Price Each, Order Date, and Purchase Address, our expected results cannot measure directly from the attributes, we need apply the adding proper columns, converting string data type into integer data type, finally apply the group by method to calculate aggregate analysis each and every calculation parts are clearly discussed in result session.

3. Result and Discussion

In this taken input data set from https://github.com/ as Sales Analysis source data, which consists sales data of 12 month various flies such as Sales_January _2019, Sales_February_2019, Sales_March_2019, Sales_April_2019, Sales_May _2019, Sales_June_2019, Sales_July_2019, Sales_August_2019, Sales_ September_2019, Sales_October_2019, Sales_November_2019, Sales_ December_2019. We need analyze and predict the sales based on the all the dataset, which need to merge into single dataset called all_data using python coding , from that merged files we need to apply our various prediction. The data set Fig.2 contains following fields like Order ID, Product, and Quantity Ordered, Price Each, Order Date, and Purchase Address. The final data set consist of all_data contains hundreds of thousands of electronics store purchases broken down by month, product type, cost, purchase address, etc.

The first phase of our work is pre-processing, which will remove all the missing values from dataset, using isnull() function find the all the missing values from data set and deleted, because all the missing always reduce the accuracy of the final expected results. Second phase is type conversion, which to convert the data either categorical to numerical or numerical to category using astype method (). Our expected prediction calculation we need to understand, which are columns are major impact on our output prediction, and either concern column values need to type conversion or not.

The Fig.3 shows the extracted data of monthly products sales from sales data set, the original data set having Order Date attribute, from this monthly sales are extracted using string functions, again same string data is converted to integer

Index	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
73891	147268	Wired Headphones	1	11.99	01/01/19 03:07	9 Lake St, New York City, NY 10001
74701	148041	USB-C Charging Cab…	1	11.95	01/01/19 03:40	760 Church St, San Francisco, CA 94016
76054	149343	Apple Airpods Headphones	1	150	01/01/19 04:56	735 5th St, New York City, NY 10001
76708	149964	AAA Batteries (4-pack)	1	2.99	01/01/19 05:53	75 Jackson St, Dallas, TX 75001
76061	149350	USB-C Charging Cab…	2	11.95	01/01/19 06:03	943 2nd St, Atlanta, GA 30301
68056	141732	iPhone	1	700	01/01/19 06:13	446 Pine St, Atlanta, GA 30301
76348	149620	Lightning Charging Cab…	1	14.95	01/01/19 06:34	338 Chestnut St, San Francisco, CA 94016
68826	142451	AAA Batteries (4-pack)	1	2.99	01/01/19 06:41	232 12th St, Boston, MA 02215
72609	146039	34in Ultrawide Mo…	1	379.99	01/01/19 07:24	53 River St, San Francisco, CA 94016
67631	141316	AAA Batteries (4-pack)	3	2.99	01/01/19 07:26	235 South St, Seattle, WA 98101
69934	143498	AA Batteries (4-pack)	3	3.84	01/01/19 07:26	428 Highland St, New York City, NY 10001
71310	144804	iPhone	1	700	01/01/19 07:29	628 Lake St, New York City, NY 10001
71311	144804	Wired Headphones	1	11.99	01/01/19 07:29	628 Lake St, New York City, NY 10001
71806	145270	Google Phone	1	600	01/01/19 07:33	392 4th St, Dallas, TX 75001
69180	142789	AAA Batteries (4-pack)	1	2.99	01/01/19 07:35	336 Spruce St, Boston, MA 02215

FIGURE 2. Combined data set of 12 Months (186850 Rows, 6 Columns)

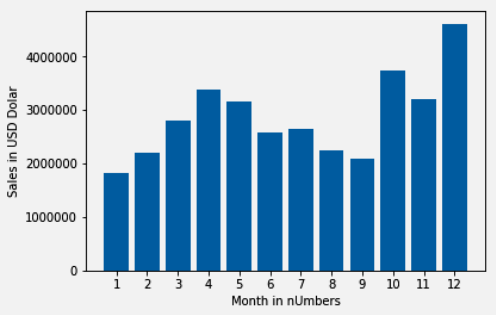

FIGURE 3. Products sales monthly report

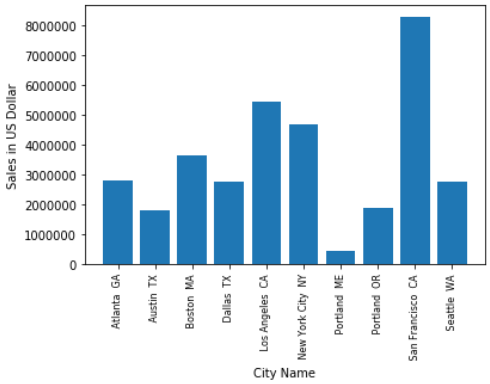

FIGURE 4. Sales of Products City wise

data type using astype() function, then finally using data preprocessing method ,eliminated all the missing values from the sales data set graphical represented as bar plot, X-axis shows the months in number from January to December and Y-axis represents the Sales in US Dollar. The as the graphical representation, every year begins the products are sales very less, then it will be increased gradually. The month of December has the maximum sales in dollar. The Fig.4 shows the sales of products city wise between X-axis name of the city and Y-axis products sales in US Dollar. The source data set does not having the city column from

data set, which is extracted Quantity Ordered, Price Each attribute from the data set. The first step both column data are converted into integer using to_numeric function in python, then sales is calculated based on the below formula.

$$Sales = QuantityOrdered * PriceEach$$

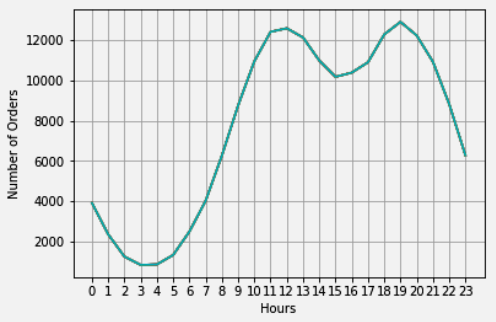

FIGURE 5. Peek Time for Advertisement (11 am or 7 Pm)

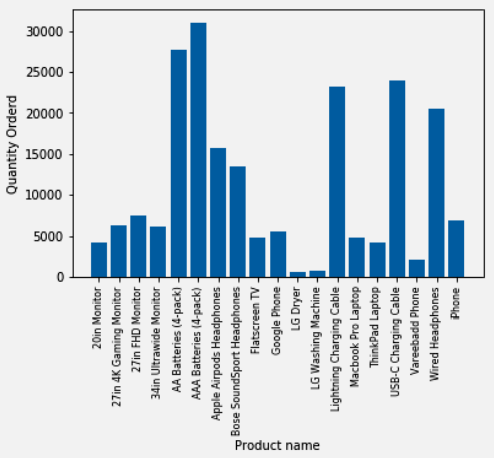

FIGURE 6. Product wise sales

The Fig.5 demonstrates the peek time for advertisement, to purchase maximum customers based on the sales data set, the X-axis represents the hours and Y-axis represents Number of orders purchased, both data are extracted from Order Date attribute. The graphs show best time for advertisement is 11 am and 7, pm are most peek time for customer visit and purchase the orders. The Fig.6 which represent the individual product wise sales with respective count, from which X-axis represents product names and Y-axis represents the quantity of items are ordered by the customer, then we need to find , reasons for maximum items are purchased by customer, based on the attributes of sales data set.

Fig. 7 represents the product name sales in between Quantity Ordered and Price of the product, In which X-axis represents the product name and having two Y-axis such as Y1-Axis is Quantity of Ordered or Number of Sales of the product and Y2-Axis shows the price of the individual products, from this group we clearlly understand AAA-Batteries are sold more which represents green color

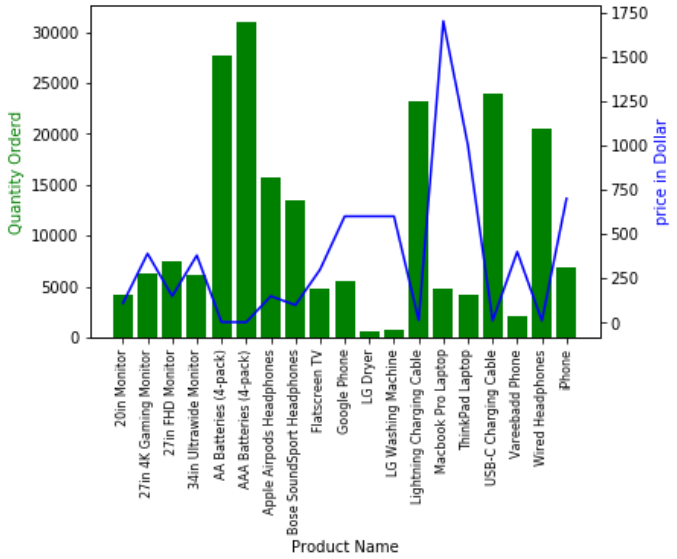

FIGURE 7. Product sales prediction based on Quantity Ordered and Price of the Product

bar chart and price is less represented blue color link so Price and quantity order are dirctly proposal.

CONCLUSION

A Sales analysis is very importatant constraint to run any business successful.The prediction of customers purchased item is playing important role in Sales Analsysis, In this proposed work subsequent task are implemented using python coding like data analysis such as best month for sales, best month for earned more,city sold more product, best time for adverisement to maximize the likelihood of customer buying product ,products are purchased by togeather finally what product are sold more resaon behind of selling more based on the Sales Dataset. All the predicted statistical analysis will be help for proper planning and keep storing of all stocks in upcoming years, easily to identify the customer expectation and increase the sales of the products.

REFERENCES

[1] K. BURLAKANTI, R. VIJAYA: *The Most Influential Factors of Consumers' Buying Pattern at Organized and Unorganized Retail Stores with Special Reference to Kakinada City, Andhra Pradesh*, Indian Journal of Marketing, **43**(1) (2013), 2013.

[2] B. SHANNAQ, K. F. SADRIWALA: *Predicting Consumer Buying Behavior Pattern using Classification Technique*, The International Information Systems Conference (iiSC), 2011.

[3] S. QAZZAFI: *Customer Buying Decision Process Towards Products*, International Journal of Scientific Research and Engineering Development, **2**(5) (2019), 130-134.

[4] P. HANSEN, H. HEINSBROEK: *Product selection and space allocation in supermarkets*, European Journal of Operational Research - **3**(6) (1979), 474-484.

[5] K. HJORT, B. LANTZ, D. ERICSSON, J. GATTORNA: *Customer segmentation based on buying and returning behaviour: Supporting Differentiated Service Delivery in Fashion E-Commerce.*, In: Pawar K.S., Rogers H., Potter A., Naim M. (eds) Developments in Logistics and Supply Chain Management. Palgrave Macmillan, London. (2016). https://doi.org/10.1057/9781137541253_14

[6] Y.-C. LEE, Y.-C. WANG, S.-C. LU, Y.-F. HSIEH, C.-H. CHIEN, S.-B. TSAI, W. DONG: *An empirical research on customer satisfaction study: a consideration of different levels of performance*, Springer Plus, **5**, Art. no. 1577, (2016). https://doi.org/10.1186/s40064-016-3208-z

[7] W. HUANG, Q. ZHANG, W. XU, H. FU, M. WANG, X. LIANG: *A Novel Trigger Model for Sales Prediction with Data Mining Techniques*, Data Science Journal, **14** (2015), 15 pages. DOI: http://doi.org/10.5334/dsj-2015-015

[8] R. KUO, K. XUE: *A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights*, Decision Support Systems, **24**(2) (1998), 105–126.

[9] G. SCHROEDER, A. KLIM, G. HEINZ, ET AL.: *System for predicting sales lift and profit of a product based on historical sales information*, U.S. Patent 7689456.

[10] H. YUAN, W. XU, M. WANG: *Can online user behavior improve the performance of sales prediction in E-commerce?*, IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, 2014, 2347-2352. doi: 10.1109/SMC.2014.6974277

[1,5]SANJIVANI COLLEGE OF ENGINEERING, SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE, MAHARASHTRA, INDIA.

Email address: 1vengicse2005@gmail.com

[2,4]SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, DEPT OF ECE, CHENNAI, INDIA.

[3]DEPARTMENT OF COMPUTER SCIENCE, BANARAS HINDU UNIVERSITY, VARANASI, INDIA.

Email address: [3]abhishek.maacindia@gmail.com

Email address: [4]yuvasivasanthi@gmail.com

Email address: 5punjabishiv@gmail.com