Data Analysis of Betelnut's Selling Dataset Using Tableau

Vikas Rahate
Department of Computer Science and Engineering
Lakshmi Narain College of Technology(R.G.P.V)
Indore, India
vikrah0921@yahoo.com

Manisha Yadav
Department of Computer Science and Engineering
Lakshmi Narain College of Technology(R.G.P.V)
Indore, India
yadavmanisha181@gmail.com

Badal Chouhan
Department of Computer Science and Engineering
Lakshmi Narain College of Technology(R.G.P.V)
Indore, India
badalchouhan295@gmail.com

Seema Narware
Department of Computer Science and Engineering
Lakshmi Narain College of Technology(R.G.P.V)
Indore, India
seemanarware@gmail.com

Khushboo Sawant
Department of Computer Science and Engineering
Lakshmi Narain College of Technology(R.G.P.V)
Indore, India
sawantkhushboo@gmail.com

Abstract— The research work aims to compress the large real-time dataset into their insights with proper procedures. It's a clear demonstration of the Data Analysis process with proper results and their interpretations which has been estimated on the basis of the dataset. This paper was created with a motto to demonstrate a clear idea about data analytics on a simple complex and larger dataset of Betelnuts selling across India in the current year 2020. The methodology explained here strictly focuses on Data Exploration, Model Development, and Data Visualization with basic knowledge of geography, culture, and tradition of our motherland India. Implementation has been done with the use of Tableau software. The implementation of this approach is done with a vision to give a clear idea of the use of data analytics at the local level in the country. Keywords - Data Analytics, Visualization, Manipulation, Interpretation, Estimation, Tableau, IDC, DMI;

I. INTRODUCTION

Data science and Data Analysis are though not newly coined terms in today's digital world and 4G era which if filled up with wholesome data. Data science has become a new trend in Data Analytics and Visualization due to the sudden increase in the use of technology and digital growth of e-platforms and electronic accessories. As it's a fact that where there is a computer, there is data, and where there is data there is information that could be further used to draw out insights. According to I.D.C., the data is grown by a factor of 300 in the last 15 years from 130 exabytes to 40000 exabytes. The main aim of Data Analytics is to draw out insights to make the decision-making process faster easier and to make it profitable. Most of the predictions in today's technical era are made using analytical tools and techniques to meet the demand of the information deficit world. Data analysis comprises data gathering, data cleaning, and using it to model the data, to get conclusions to improve the productivity and growth of the respective firm, industry, or business organization. It has been quoted that information is the oil of the 21st century and analytics is the combustion engine. Therefore, data must be processed to get the information refined (oil) and enter the arena of data analysis. After entering the data science field, some computational tools are needed to interpret and tackle the huge amount of complex data because normal human is not able to manage huge data and tackle the same. The problem which is tackled here and of which the solution is provided is the lack of an analytical approach in the current time at the local level of marketing and production. Most of the time this approach is not followed which leads to the wrong implementation of strategy and may result in loss of both financial as well as social credit

II. LITERATURE REVIEW

Baeder F. et al. [1] discussed Description logic is a family of knowledge representation formalism that goes down from semantic networks and frames via the system Kl-one. During the last decade, it has been shown that the important reasoning problems in a variety of description logic can be distinguished using tableau-like algorithms. This is very unsurprising since description logic has turned out to be closely related to propositional modal logic and logic of programs (such as propositional dynamic logic), for which tableau procedures have been quite successful. Akhtar et al [4] have investigated Data analytics and visualization using Tableau utilitarian for COVID-19 (Coronavirus). A huge opportunity provided to use this data and to turn these opportunities into reality. At this time spread of COVID-19, peoples must be made available with reliable, trustworthy information. At this time tableau played an important role, It's a very powerful tool for visualizing datasets very easily. For any type of data, Tableau is a very useful tool to generate an output that is more understandable and presentable in the visualization techniques. It also includes data blending, real-time reporting, and collaboration of data. In April 2014, Grolemund G. and Wickham H. [5] investigated A Cognitive Interpretation Data Analysis, the current paper has explained the data analysis process in the form of a scientific model. In this, it has been argued that data

analysis is primarily a procedure to develop understanding because it includes several processes. Data analysis tasks closely resemble the cognitive process known as sense-making. it has been demonstrated here that how data analysis is a sensible task adapted to use quantitative data. The problems explained form a useful organizing model for the data analysis task while allowing methods to remain flexible and situation dependent. The insights of this model are especially helpful for analysts, statisticians, and teachers of data analysis. In 1994, Susan Spiggle [6] has investigated on Analysis and Interpretation of Qualitative Data in Consumer Research and explained the work done by him in the field of qualitative data analysis. Herein it truly claimed about the contradiction of two words which are data analysis and interpretation. Many consumer researchers describe how they tackled the data by stating that subcategories, their perspectives on this has been clearly explained. Here, it has also been explained categorization, integration, refutation concerning data analysis. Also explained that refutation-provide a means for managing qualitative data for the purpose of analysis and interpretation. The focus here has been on the analytical procedure, rather than interpretations' springing from it.

III. Methodology

1. Techniques used

The main methodology and process are based on the following key techniques.

1.1 Data Exploration

Understanding characteristics of given fields in the given data which are variable distributions of, if the dataset is giving some indication, or data is invalid or what. Exploring the interrelation between real-time values and variables and also further, transforming data so that it becomes eligible for analysis. It will always include checking the datatype of variables whether the values are indicating to appropriate aggregate level or not. If they are not doing so, then creating additional variables by extraction and manipulation of these variables to get new simplified variables. Now newly created variables can help out to get insights. Limitations and exclusions for data are keys to data exploration.

1.2 Model Development

Determining a hypothesis related to the business questions that can be answered with the data. Herein the dataset used is betelnuts growing and selling data as per state and district-wise selling of betelnuts in respected markets after proper growth and plucking. The production of statistical graphs and charts is done here to reach specific conclusions which are total indications of what the data is symbolizing. The creation of calculated fields (a month in our dataset) also helps to extract the real trends in the selling of a product in the local market on monthly basis. Other variables or fields can be manipulated according to the need of trend or explored content, which may be an indicator of whether the commodity sold is at a profitable price or going in loss. Modeling the data in Tableau actually means feeding the data to make it suitable to feed inside the workbook using the drag and drop system to draw graphs and charts.

1.3 Interpretation and Result Reporting

Visualizing and presenting the insights. This involves the interpretation of significant variables which can help in predicting the future perspective of the current analysis. A business perspective is always concerned while assigning priority or importance to any variable or task. After interpreting, proper reporting of conclusions is done concerning conclusions from the obtained graph or chart or any other statistical representation.

2. Tools and Data-set used

The tool has used here is Tableau [1] which is software based on business intelligence and data visualization used for data analytics [3]. A tableau is a tool developed by Salesforce which is an American software company. It provides multiple data connection facilities like direct inclusion of dataset or giving the server link of a specified dataset. The dataset we've used is a publicly accessible dataset, approved and released through the Ministry of Agriculture and Farmer's welfare by the Directorate of Marketing and Inspection (DMI), Government of India.

The dataset is related to variety-wise daily market prices of Betelnuts sold across the country in the current fiscal year. It's released under National Data Sharing and Accessibility Policy (NDSAP) either for research work or public use. Betelnut is one of the finest crops of Western Ghats as well as the North-Eastern region of Meghalaya and Tripura. The dataset is modified by adding certain columns under model development.

3. Concept of Research

To carry out analysis from the dataset using Tableau and get some interesting insights which are to be followed according to given data. To evaluate domains and get interpretations about data-fields. Here are some graphical and statistical representations to get a clear idea of what the data is indicating and about what is to be concerned further in near future. The dataset which is used here is cleaned by likely removing some null fields, abstract data, or null values using MS-Excel formulae and our mind set to get the data visualized such that it could be interpreted more easily rather than making it complex to reach to a specific decision.

IV. Results and Discussion

After analyzing the data, here are some interpretations and results which are clear indicator of what data is conveying on the basis of which predictions and estimations for near future are to be made so that it will help and benefit us.

Following are some results and clarifications obtained from our small analysis.

1. The shaded region in the above map indicates the states producing Betelnuts in India.

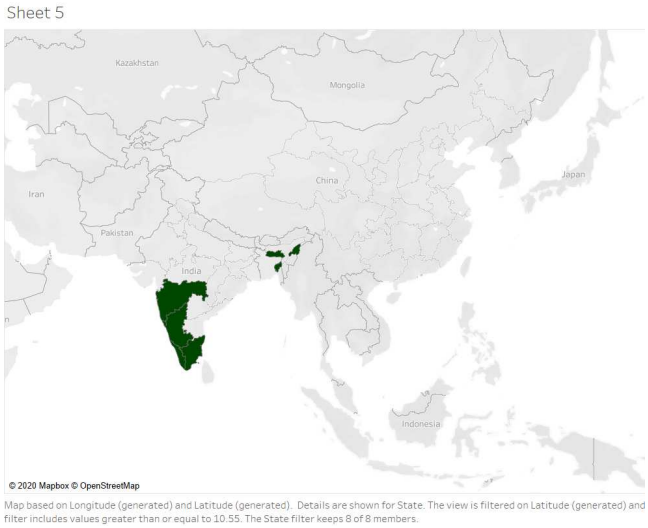

Fig. 1 Indication of geographical regions covered by Betelnut cultivation in the country.

Betelnut mostly grows in tropical and sub-tropical regions. It is a profitable plantation crop that is commercialized for maximizing returns. The main reason for cultivation in the above-shaded regions is well-distributed rainfall and well-drained soil with a low water table. The above regions mostly southern India and northeastern part are having laterite, loamy (red), and alluvial soil, which is suitable for its plantation. The temperature range required is usually 15°C-37°C.

2. Bar graph for Month of Arrival for selling versus Average of Maximum price per month for Betelnuts sold in 2020

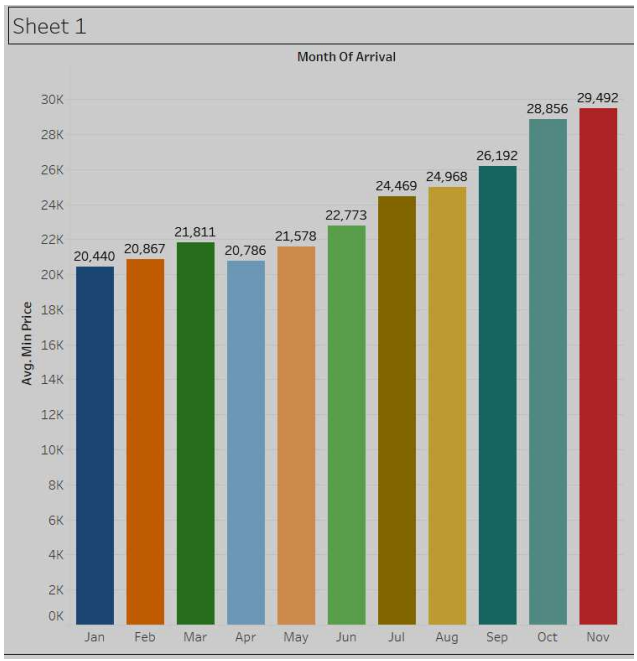

Fig. 2 Bar graph of Month of arrival versus average maximum price

From this graph, it could be concluded that the average maximum price is highest for the month of November with an average difference of Rs 9693 between the starting month of January and the end of November, which is a huge value for an average farmer in India. Now, this could be inferred and predicted as per the current production year that the rates are constantly increasing (except in the month of March) towards the end of the year. So this would ultimately be beneficial and profitable to sell the crop produce final months of the year or winter season. One reason for this may be the storage of same for the season of marriages in the coming summer season and also due to increase in religious occasions like Navratri, Durga Puja and Deepawali throughout the country, because Betelnut (supari) has its usual traditional importance in Indian rituals and traditional events as well as festivals.

2. The 2nd result obtained is a line graph between the average maximum price and the average of the maximum difference.

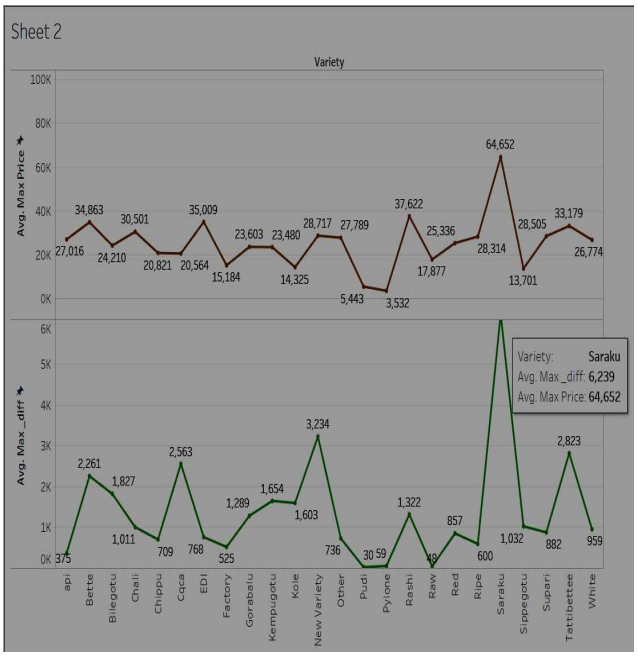

Fig. 3 Line graph between average maximum price and average of maximum differences.

From this, it can be interpreted that raw and pudi variety Betelnut is having least maximum price difference and highest profit is obtained from it as far as other varieties are concerned so this a profitable variety because of least deviation from the median value which is 985, with reference to modal price. So, farming of raw and pudi must be emphasized for cultivation to maximize the profit whereas saruku must be avoided due to less profit and demand. The least profitable variety is saruku, because of the highest deviation from the median value which is Rs. 26,055. The peak obtained for the difference of maximum difference is obtained in the case of saruku which is Rs 6239.

3.The third obtained result is from state-wise Bar graph between average maximum price and average modal price.

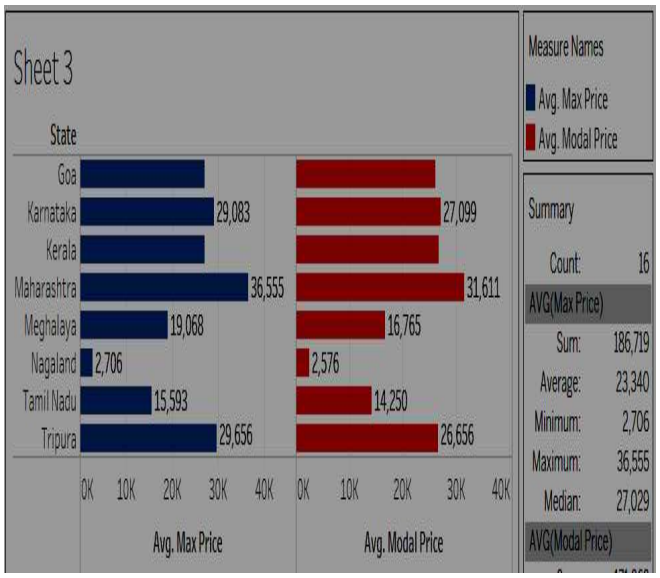

Fig. 4 Bar graph between average maximum price and average modal price.

From this bar chart, it could be estimated that Maharashtra is the state with the highest average market value for Betelnuts. So, it is advisable to farmers around Maharashtra that it is the state with the highest profit value. The neighboring state must be advised to sell their crop produce in Maharashtra only if their transportation value is affordable. In the meantime, it must be kept in mind that Mumbai is the only selling market available in the dataset from Maharashtra. Also, it is having a higher crop rate than other states with an average difference of Rs 13,215. Before advising it must be kept in mind that one of the issues would be accommodation cost to reach out to another state. For this Government of India with Ministry of Railways and Agriculture have given the facility of compensated train transportation for farmer's crops produce.

4. The result obtained is a line graph between markets of various states concerning the model price specified by market authorities. Though most of the markets are over the median value of the maximum price the market with the highest modal price must be prioritized as far as income and profit are concerned. This could be predicted that Bangalore is the city with the highest average model price for selling Betelnuts. So Bangalore in Karnataka with average Rs 49,591 then Mumbai in Maharashtra with Rs. 31,611, Manjeswaram in Kerala with Rs. 30,755, Pernaam in Goa with Rs. 28,960, Masli from Tripura with an average of Rs. 26,656, Thondamathur in Tamil-Nadu with an average of Rs. 21,842, Shillong in Meghalaya with Rs. 16,765, and Julukie in Nagaland with Rs. 6300 are the state-wise top markets for selling.

Though Data analysis and Tableau helped a lot in predicting and getting insights and conclusions from the complex dataset.

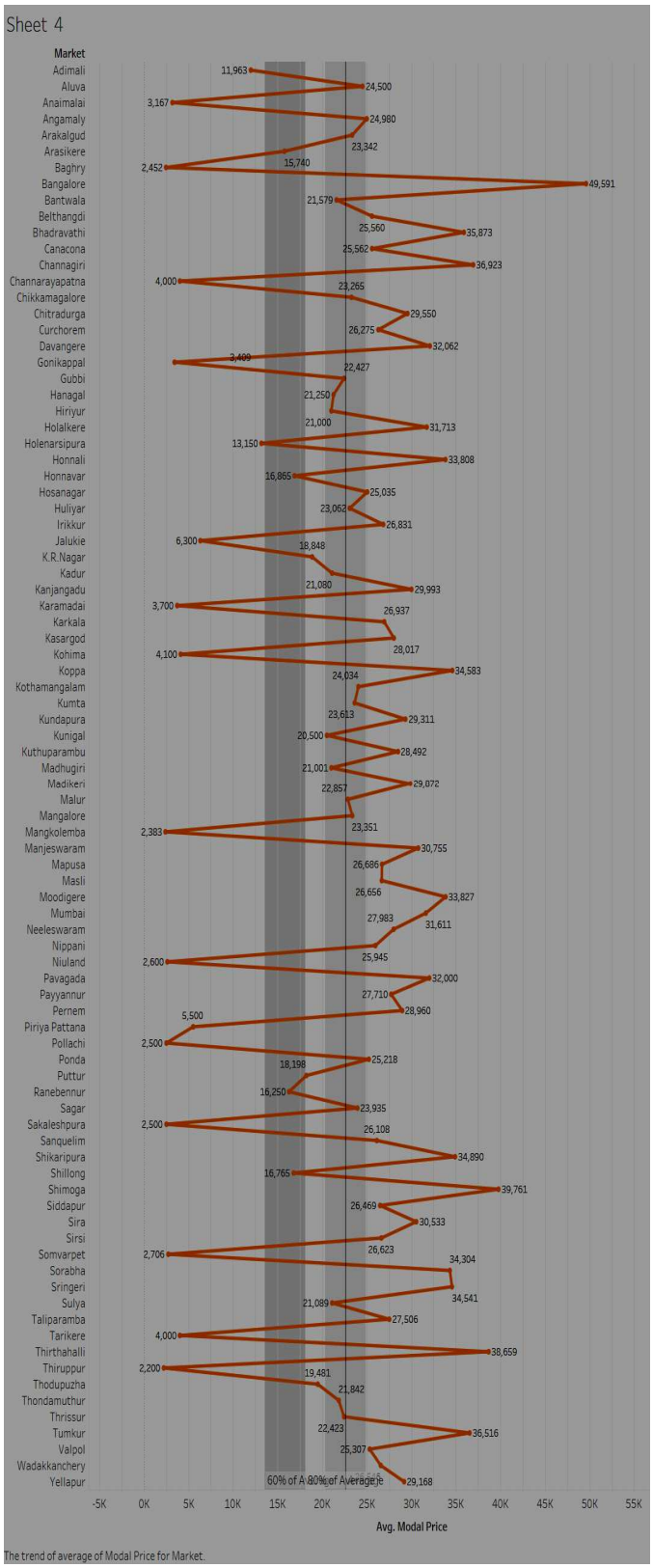

Fig. 5 Line graph between markets versus average model price.

Tableau proved to be useful for faster and easier analysis for large as well as small datasets. It also helps in faster interpretation using easier methods like summary, highlighters and parameters also show mark labels helped in getting numerical interpretations of respective graphs.

V. CONCLUSION

From the discussion and results analyzed above, it can be concluded that no doubt Tableau played a vital role and proved to be a useful analytical tool for analysis, interpretation, and visualization [3] of data with a simple approach of data gathering, cleaning, modeling and result from analysis. In our research work, we've analyzed the data from the Ministry of Agricultural Development and Farmer's welfare. From this data we've concluded that (1) the best time to sell betelnuts is last quarter of the year to get maximum price (2) pudi and raw betel nut are the most profitable betelnut varieties as far as maximum price is concerned, whereas suruku is least demanding and hence, less profitable (3) average maximum selling price is highest for Maharashtra followed by Karnataka and Tripura (4)some state-wise cities with the highest demand and hence higher price are already being discussed in the result.

VI. FUTURE WORK

The currently proposed work is for analyzing the basic trends in Data Analytics the future work would be to analyze the dataset's combining the past 10 -15 years' production and selling analysis including gulf export and foreign transfer concerning demand, category, variety of Betelnut or any other crop produce and dimension of the respected country in terms of importing of Betelnuts. The future work will definitely lead to historic advancements in the field of Agricultural production. Interpretation of such huge data with proper command over the results will ultimately remove the problem which is explained here that is lack of analytical approach at the local level of marketing and production. This will ultimately help farmers and local traders to boost their financial growth.

VII. ACKNOWLEDGMENT

We would like to express our sincere gratitude to advisor **Dr. Harish Patidar** and **Dr. Chetan Choudhary** for the continuous support and also for motivation, enthusiasm and immense knowledge. Their guidance helped all the time for study and research. We would like to acknowledge the contribution of all the people who have helped in reviewing this paper.

VII. REFERENCES

1. Baader F, Sattler U, an overview of Tableau Algorithm for Description Logic

2. Stolte C., Tang D., and Hanrahan, P. 2008. Polaris: a system for query, analysis, and visualization of multidimensional databases. Communication. ACM 51, 11 (Nov. 2008), 75--84.

3.Government of India, National Informatics Centre, DMI

4.Akhtar, Nikhat & Tabassum, Nazia & Perwej, Dr. Asif & Perwej, Dr. Yusuf. (2020). Data analytics and visualization using Tableau utilitarian for COVID-19 (Coronavirus). Global Journal of Engineering and Technology Advances. Volume 3. Page 28-50. 10.30574/gjeta.2020.3.2.0029.

5.Grolemund, G. and Wickham, H. (2014), A Cognitive Interpretation of Data Analysis. International Statistical Review, 82: 184-204.

6.Susan Spiggle, Analysis and Interpretation of Qualitative Data in Consumer Research, Journal of Consumer Research, Volume 21, Issue 3, December 1994, Pages 491–503,

7. Jamie Hoelscher, Amanda Mortimer, Using Tableau to visualize data and drive decision-making, Journal of Accounting Education, Volume 44,2018, Pages 49-59, ISSN 0748-5751,

8. Faten Hamad, Hussam Fakhuri, Sinaria Abdel Jabbar. (2020) Big Data Opportunities and Challenges for Analytics Strategies in Jordanian Academic Libraries. *New Review of Academic Librarianship* 0:0, pages 1-24.

9.Lengere, Monica, 2020/02/25, CHAPTER FOUR DATA ANALYSIS INTERPRETATION 4.1 Data analysis, Presentation and Interpretation Plan

10.Mons, Barend. (2018). Data cycle step 6: Data analysis, interpretation. 10.1201/9781315380711-7.

11.Sarah Anne Murphy (2013) Data Visualization and Rapid Analytics: Applying Tableau Desktop to Support Library Decision-Making, Journal of Web Librarianship, 7:4, 465-476, DOI: 10.1080/19322909.2013.825148

12. Suresh K. Peddoju, Himanshu Upadhyay. 2020. Evaluation of IoT Data Visualization Tools and Techniques. Data Visualization, pages 115-139.

13.Chapman, Joyce and Lown, Cory. 2010. Practical Ways to Promote and Support Collaborative Data Analysis Projects. *code4lib*,12: 1–10.

14. Tableau Software. 2020.3 Technical Specifications—Tableau Desktop. Accessed: November 2020.

15. Few, Stephen. 2009a. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, Oakland, CA: Analytics Press. [Google Scholar]