

DATA ANALYTICS AND VISUALIZATION THROUGH R PROGRAMMING

C. Baby¹, Mr. S. Venkatesulu²

¹Reg.no:1116105, MCA VI semester, Department of MCA, Sree Vidyanikethan Institute of Management, S.V.University, Tirupati, Andhra Pradesh

²Assistant Professor, Department of MCA, Sree Vidyanikethan Institute of Management, A. Rangampeta, Tirupati

-----***-----

Abstract:- Analytics refers ability to collect and use data that generates data internally and externally. It is a fact based decision making. The current trends and advanced techniques of the information technology defines that with a very short period the data becomes a most required element in and around of the premises. When we have a data we can start with the current techniques. The applications of Data Analytics are implementing in all the domains such as Information technology, Health care, Communication, share market analysis, political parties (Making decisions to speech) and all other domains we can apply data analytics for better decision making.

Key Words: Data, GUI, Dash Boards, Plotting, Geographical Representation.

INTRODUCTION:

The data is one of the most powerful elements for analytics and visualization. The advanced techniques and other details such as debit / credit cards. Later on the retailer can analyze to understand the business better.

Techniques such as visualization, framing, data analysis helps to analyze the data for the further reference. The application domains such as healthcare, communication, market analysis, decision making is also supported by Data Analysis. The data can be collected from various sources [1]. The various sources of data is

Sources of data for data analytics:

- Whats app
- Twitter,
- google maps,
- social media,
- you tube face book

Example of data collection:

once an user visit the shop and do shopping, the shopping details are going to store in the computer such as How much an user is bought, What are the items he bought, Age, phone number

How much data a single transaction can generate:

1. Email 2.9 millions
2. House hold data consumption: 375
3. 20 Hours of videos uploaded into the facebook
4. 24 peta bites of data every single day on google
5. 50 millions a twits per day
6. 700 million minutes spend on each month
7. 1.3 hexa bytes of data sent and receive via mobile internet
8. 72.9 products ordered on Amazon every day.

DATA VISUALIZATION:

Data visualization is one of the advanced tools for data visual and understanding purpose [2]. It provides better way of understanding the data. The reports that are generated may be represented in the form of pie charts, bar graphs and line graphs. The traditional methods does not provide better understanding and the values will not be visible clearly. All these difficulties are overcome with the implementation of data visualization. The text based reports will take lot of lot of time to analyze and the lengthy text leads to confusion [3]. So we need a tool for better data analysis and visualization purpose.

The visualization tool such as data visualization was designed with the advanced data display methods so that the data can be visualize in more understand format. The common person will also be able to understand the data. The data displayed in more stylish and understandable format. Few data visualization methods are

- Sparklines
- Heat maps
- Geographic's maps,
- Infograph etc

The images are displayed in variety image formats such as Interactive, Manipulate or data drilling. he data querying and analysis will also be indicators by the developers. The users can utilize the features of the data visualization.

The features of data visualization:

- ❖ Query based data retrieving
- ❖ Perfect data analysis
- ❖ Allow the user for data manipulation
- ❖ Predefined conditions may be implemented.

LITERATURE REVIEW:

In the year 1160 first documented version of data visualization has been identified, the document with *Turin Papyrus Map* illustrates the various distribution according to geological resources and it also provides other information such querying of those resources. Physical quantities and geographic space representation was developed in the 16th century [4]. In the 20th century Jacques Bertin implemented quantitative graphs to represent information. "Rene Descartes and Pierre de Fermat developed analytic geometry and 2-D co ordinate system which heavily defined the practical methods for displaying and calculating values [5]. IDC implemented data visualization to generate the market share (predicted) from 2015-2020. The visualization, as shown in the figure 1.0.

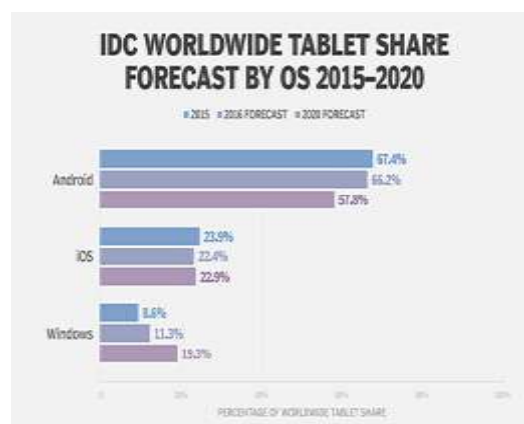


Figure 1: Forecasting Visualization

The data visualization advantages:

- Displaying the data
- Data displayed through substance rather than methodology.
- The data displayed in graphic design.
- Implementing large data sets
- Different pieces of data will be will be compare and provide the consolidated data.
- Close integration is possible with verbal descriptions on data set.

ABOUT R PROGRAMMING LANGUAGE:

One of the most popular programming languages in the 4th generation is R Programming [6]. It contains very light weight environment commonly implemented with statistical computing, Analysis of Data and implementing scientific research. The data analysis, prediction and market retrievers can implement R Programming for data derivations. The R Programming is contains easy syntax and easy-to-use interface [7][8]. The sample data visualization through R Programming as shown in the figure 2.0

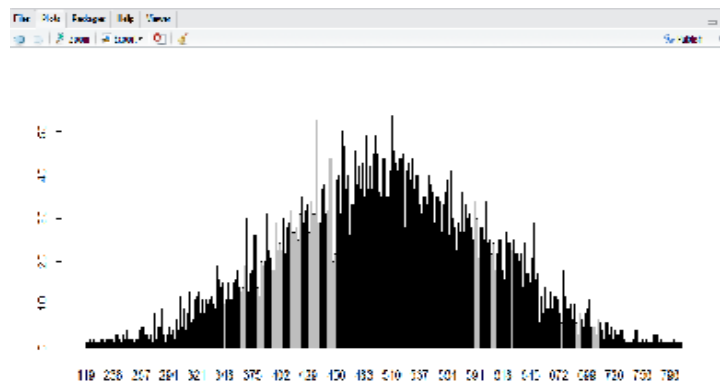


Figure 2.0: Data visualization

Stack market analysis and data visualization through R Programming:

The data was collected for various stocks for the year 2017 (from January to December) [9]. The R Programming was implemented to analyze and visualize the data [10]. The implementation process as shown in figure 3.0

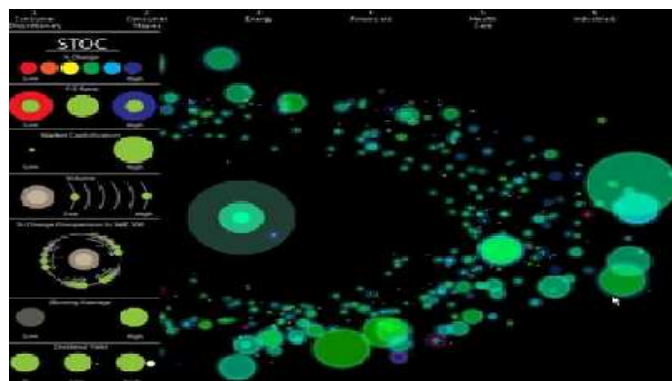


Figure 3: Stock Market Analysis for the year 2017

The data visualization based on the economy of the year 2017 as shown in figure 4.0



Figure 4: Economy Visualizaiton

CONCLUSION:

Data is one most powerful tool for analysis, prediction and forecasting. Now the entire world is completely focused on data analytics and visualization. The visualization helps to design dashboards and presenting the information in the graphical format. The R Programming helps to represent the data in GUI format. The TABLEAU software is help much suitable to generate the data visualization objectives.

REFERENCES:

1. Shaun Bangay, A system for the visualization of Multi-dimensional data, in "Visual Data Exploration and Analysis V". (TA 1505 Pse 3298)
2. Matthew Ward, Overview of Data Visualization, from www.cs.wpi.edu
3. Alfred Inselberg, "Multidimensional Detective", in "IEEE Symposium on Information Visualization 1997". (TK 7882 Inf. Iv3)
4. Finn Årup Nielsen , "Visualization and Analysis of 3D Functional Brain Images", from <http://hendrix.imm.dtu.dk/staff/students/fnielsen/thesis/finn/finn.html>
5. Pak Chung Wong, R.Daniel Bergeron, Gergory M. Nielson, Hans Hagen, Heinrich Muller, Scientific Visualization, over view, methodologies, techniques.
6. Kieran Healy, *Data Visualization: A Practical Introduction*. Princeton University Press. , First Edition.
7. Claus, O Wilke, *Fundamentals of Data Visualizaiothn*, Oreilly publications, 2014
8. Roger D. Ping, *R Programming for Data Science*, Leon Pub publications, 2015
9. Graham Williams(2009): *Rattle: A Data Mining GUI for R*, The R Journal Vol. 1/2, December 2009 ISSN 2073-4859.
10. Hadley Wickham, *R Programming Packages*, Orilley Publicaitons, 2015