Data Analysis and its Importance

Manpreet Kaur Bhatia

Institute of Innovation, Technology and Management, Guru Gobind Singh Indraprastha University, New Delhi, India-110058

Abstract— The data available is growing at an exponential rate. The increase in data in itself is a minor problem, but the percentage of unstructured data in the overall data volume is what is concerning all. so it becomes a basic necessity to discover ways to process and transform complex, unstructured, or large amounts of data—into meaningful insights, This brief outline of data analysis will help us understand what is data analysis, the value it holds in many industries worldwide and how majority of the organizations in various sectors bank on data analysis to survive the ongoing market race. This paper maintains its focus on explaining the basic procedures followed in obtaining something immensely useful from the available disorganized facts and figures by analyzing them. Also discussed briefly are its applications in areas such as management, retail, healthcare, education and so on. This paper highlights important concepts of data analysis.

Keywords— Anomalies, data analysis, decision-making, productivity enhancement, unstructured data

I. INTRODUCTION

Data are now woven into every sector and function in the global economy, and, like other essential factors of production such as hard assets and human capital, much of modern economic activity simply could not take place without them. The concept of Data Analysis — large pools of data that can be brought together and analyzed to discuss patterns and make better decisions — will become the basis of competition and growth for individual firms, enhancing productivity and creating significant value for the world economy by reducing waste and increasing the quality of products and services. In fact, no business can survive without analyzing available data. Consider the following situations:

- A pharmaceutical company is performing trials on number of patients to test its new drug to fight a certain disease.
- A company wants to launch new variant of its existing line of perfume. It wants to carry out the survey analysis and arrive at some meaningful conclusion.
- Sales director of a company knows that there is something wrong with one of its successful products, however hasn't yet carried out any market research data analysis. How and what does he conclude?

These situations are indicative enough to conclude that data analysis is the lifeline of any business. Whether one wants to arrive at some marketing decisions or fine-tune new product launch strategy, data analysis is the key to all the problems.

Inference, specifically decision making and prediction, is centuries old and plays a very important role in our lives. Each of us faces daily personal decisions and situations that require predictions concerning the future. Many individuals tend to feel that their own built-in inference-making equipment is quite good. However, experience suggests that most people

are incapable of utilizing large amounts of data, mentally weighing each bit of relevant information, and arriving at a good inference. Thus, scanning of data and making concrete inferences based on the data available becomes a subject of supreme interest.

II. WHAT IS DATA ANALYSIS?

The process of scrutinizing raw data with the purpose of drawing conclusion about that information is called "Data Analysis". The main aim of Data Analysis is to convert the available cluttered data into a format which is easy to understand, more legible, conclusive and which supports the mechanism of decision-making.

The whole process of data analysis begins with the question "what is to be measured?" The answers to these questions gives a researcher a clear idea about the main motive that the analysis should address. The potential of data analysis is in its ability to solve business problems and provide new opportunities. So in order to gain the most from your investments, focusing on the questions which support your ideas becomes a priority. Equally important are the situations which might provide hindrances in the smooth working of any organization or business. This helps in forming an objective of the analysis and justifying its need.

Arguably, the most critical aspect of a analysis is its quality or validity. An analysis is considered valid if it measures what it was created to measure. An analysis is generally considered profitable if the scores it produces help individuals make accurate inferences about a particular characteristic, trait, or attribute intrinsic.

Once the purpose behind the analysis is clear, the next step is to collect the data which is to be analyzed.

This includes the process of identifying the audiences which is to be targeted and the methodology which is to be incorporated in order to extract the desired information from the respondents.

The process of data analysis is initiated only after all the required data has been collected. Hereafter, it becomes necessary to make the data full of errors and dropping all of the unwanted information which might be present in the data. It is advisable to collect data in its raw, detailed form and then record patterns. This enables the researcher later to analyze the original data in different ways, perhaps to answer deeper questions than originally conceived.

If researchers have collected documents from participants, such as logs, journals, diaries, memos, and letters, these can also be analyzed as raw data. Similarly, official documents of an organization can be subjected to analysis [9].

The next phase comprises of scrutinizing and displaying the data. In order to condense the essence of the data it is examined thoroughly. One of the major hallmarks of

conducting research is that data are analyzed continually, throughout the study, from conceptualization through the entire data collection phase and into the interpretation and writing phases. Characteristics of the data may be described and explored by drawings and charts, doing cross tabulations.Tables are used to organize the data at one place. Relevant column and row heading facilitates finding information quickly. Once a data is organized in a tabular form spotting beneficial trends and anomalies becomes effortless. The main asset of tabular representation is its versatility and so it accommodates both qualitative or quantitative data and even the combination of both.

Further, the data is assayed using various statistical tools. The measures of central tendency indicate what is typical of the average subject. Measures of variation indicate the distribution of data around the centre and correlation refers to the degree to which two variable move in sync with one another. Some statistical test like Hypothesis testing, estimations and modeling make inferences about population. They are used to determine whether it is possible to draw general conclusions about a population, or predictions about the future based on the collected data. Advanced modelling techniques may eventually be used to build sophisticated explanations of how the data addresses the original question. But many research projects would never need to go that far; the question would be answered by simple descriptive statistics [6].

The results calculated using statistical testing are then represented graphically. Tabulation is a good method of condensing and representing data in a readily understandable form, but many people have no taste for figures. They would prefer a way of representation where figures could be avoided. This purpose is achieved by the presentation of statistical data in a visual form. The visual display of statistical data in the form of points, lines, areas and other geometrical forms and symbols, is the most general terms known as Graphical Representation. Statistical data can be studied with this method without going through figures, presented in the form of tables.Graphs are visual means of representing data. They allow complex data to be represented in a way that is easier to spot trends by eye. Graphs are perceived by our visual system. They give numbers shape and memorably. And so graphical displays are powerful tool for persuasion and contribute in a major way in the process of decision making.

Another advantage of graphical representation of data is that it is much neater and more organized than just having a bunch of figures thrown around. Also, engineers and scientists try to incorporate graphs wherever possible as it is a professional way of working with figures [4], [5].

III. BENEFITS OF DATA ANALYSIS IN VARIOUS SECTORS

International performance improvement guru and businessman H. James Harrington has rightly quoted that "Measurement is the first step that needs to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it. "

Thus measurement of any flaws or strategies which impede a company from reaching its full potential becomes a highly preferred task .So perhaps it becomes inevitable on the part of the governing body of any organization or business to pay good considerations to the data analyzed as it depicts the path with least resistance to success. It helps any business or organization identify performance problems that require some sort of action. Sophisticated analysis of data can substantially improve decision making, minimize risks, and unearth valuable insights that would otherwise remain hidden. Data analysis unlocks significant values by making certain facts and information transparent and recognizable. It allows the identification of important and often mission- critical trends.

Financial institutions can quickly find that data analysis is adept at indentifying fraud before it becomes widespread, preventing further damage. Governments have turned to data analysis to increase their security and combat outside cyber threats.

The healthcare industry uses data analysis to improve patient care and discover better ways to manage resources and personnel. The data analysis software available today, are of major benefits to the healthcare sector. As information becomes increasingly available, comparable and unambiguous, patients will also be empowered and more involved in their own treatment through online health applications, which can integrate patient information with their health records and make it available to clinicians. [3]

Further, with human genome mapping and Data Analysis tools, it will soon be commonplace for everyone to have their genes mapped as part of their medical record. This brings medicine closer than ever to finding the genetic determinants that cause a disease and developing drugs expressly tailored to treat those causes — in other words, personalized medicine [1].

Data analysis on students behaviour can provide the concerned authority with important insights, such as if a student requires more attention, the class understanding of a topic is not clear, or if the course has to be modified [7].

Telecommunications companies and other utilize data analysis to prevent customer churn which also planning the best ways to optimize new and existing wireless networks. Markets have quite a few easy to utilize data. One involves sentiment analysis, where marketers can collect data on how customers fell about certain products and services by analyzing what consumers post on social media.

Data analytics allows you to personalize the content or look and feel of your website in real time to suit each consumer entering your website, depending on, for instance, their sex, nationality or from where they ended up on your site. The best-known example is probably offering tailored recommendations: Amazon and Flipkart are using real-time, item-based filtering to fuel its 'Frequently bought together' and 'Customers who bought this item also bought' features. And the approach works as Amazon and Flipkart have recorded an increased revenue via this method.

IV. CONCLUSION

The world is becoming data driven. Each and every decision is now based on taken based on the data available.

The concept of data analysis- large pool of data can be brought together and analyzed to discuss patterns and make better decisions – will soon become the basis of competition and growth for individual firms, enhancing productivity and increasing the quality of products and services.

Vast amounts of data are being generated in structured and unstructured form. And so today companies are looking for new ways to transform their business via data to get insight to make the best possible decisions. So data and its analysis has not only revolutionized the data world but it took the employee analyst to a different level as well.

Companies are hiring the best data analysts available in the market to maximize their profit. Data analysis is able to process and store that data and so organizations are becoming more flexible and more open to data analysis [8].

`Note that one of the biggest problems data analysts very often have is that the data they get is not suited to answer the questions they are asked. For instance, if we were supposed to use the data in our customer database to find out how to differentiate Asian shopping behaviour from European, we might have a very hard time since we are assuming that we used a nice, representative sample of all different types of European shoppers to generate the data—very often this is not the case, and the data itself is already biased and will bias our analysis results [2].

REFERENCES

[1] M. A. Alkhatib, "Analysis of research in healthcare data analytics," *Australasian Conference on Information Systems*, Sydney, pp. 1-16, 2015.

[2] M. R. Berthold, *Guide to Intelligent Data Analysis*, Texts in Computer Science 42,© Springer-Verlag London Limited 2010

[3] W. S. Cleveland, *The Elements of Graphing Data*, Pacific Grove, CA: Wadsworth & Advanced Book Program, 1985.

[4] W. S. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *Journal of the American Statistical* Association, vol. 79, issue 387, pp. 531–540, 1984.

[5] A. Lacey and D. Luff, "Qualitative data analysis," *Trent Focus*, 2001.

[6] A. G. Picciano, "The evolution of big data and learning analytics in American higher education," *Journal of Asynchronous Learning Networks*, vol. 16, issue 3, pp. 9-20, 2012

[7] Sabia and K. Sheetal, "Applications of big data: Current status and future scope," *International Journal on Advanced Computer Theory and Engineering (IJACTE)*, vol. 3, issue 5, pp. 25-29, 2014

[8] P. Saxena, "Application of statistical techniques in market research: A sample survey," *International Journal of Applied Engineering Research*, *Dindigul*, vol. 2, no 1, pp. 163-171, 2011.