# An Impact of Empirical Data Analysis in the World of Business Environment

## Merla Swetha[1], Naresh E[2], Santosh Parakh[3]

1.  School of Management Studies, Reva University, Bangalore, Rukmini Knowledge Park, Kattigenahalli, SH 104, Srinivasa Nagar, Bengaluru, Karnataka 560064, India
2.  Department of ISE, M S Ramaiah Institute of Technology, Bangalore, MSRIT Post, M S Ramaiah Nagar, MSR Nagar, Bengaluru, Karnataka 560054, India
3.  Department of MCA, Vidya Pratishthan's Institute of Information Technology, Vidyanagari, Vidya Nagari, Baramati, Maharashtra 413133, India

ARTICLE INFO

Corresponding author:
merlaswetha@gmail.com

ABSTRACT

Data Analysis plays a key role in all the fields (in Production and Operations, Finance, Marketing, Human Resource Analysis and Management) for successful market share. Data science can raise the value of any business who can effectively utilize their data. Data should be at the fingertips of each organization's leader to make strategic decisions. The proposed work presents heterogeneous types of data analysis techniques that are useful for changing the needs of the business. This study showcases how effective data analysis is carried out both qualitatively and quantitatively. The utilization of data analysis is to identify and distinguish the relationship in data and the trends among the factors which represent the data. It plays a vital role in making a business decision more effectively. With the aid of different data analysis techniques, data scientists can perform many operations on datasets that are useful for business organizations and it can also, help in many industrial applications.

SARI PATI

*Analisis Data memainkan peran kunci dalam semua bidang (Produksi dan Operasi, Keuangan, Pemasaran, Analisis dan Manajemen Sumber Daya Manusia) untuk pangsa pasar yang sukses. Ilmu data (data science) dapat meningkatkan nilai bisnis apa pun yang dapat secara efektif memanfaatkan data mereka. Data harus berada di ujung jari pemimpin masing-masing organisasi untuk membuat keputusan strategis. Pekerjaan yang diusulkan menyajikan jenis teknik analisis data heterogen yang berguna untuk mengubah kebutuhan bisnis. Studi ini menunjukkan bagaimana analisis data menjadi efektif jika dilakukan baik secara kualitatif maupun kuantitatif. Pemanfaatan analisis data adalah untuk mengidentifikasi dan membedakan hubungan dalam data dan tren di antara faktor-faktor yang mewakili data. Ini memainkan peran penting dalam membuat keputusan bisnis lebih efektif. Dengan bantuan teknik analisis data yang berbeda, ilmuwan data dapat melakukan banyak operasi pada dataset yang berguna untuk organisasi bisnis dan juga dapat membantu dalam banyak penerapannya di industri.*

## INTRODUCTION

Data science gives significant information dependent on a lot of multifaceted or large data. A data fact, or facts obsessed science, joins a variety of fields of work in insights and calculation to decipher data for basic leadership purposes in running the business successfully. Data science utilizes logical strategies, procedures, calculations; structure to eradicate facts Information should be at the fingertips of each organization's leader. Information helps to take noteworthy decisions those are key imperative practices for any business.

Data science itself is advancing as a discipline in the modern world. Data is power in business management, and data is the fuel that makes this force. It is assessed that by 2020, information volume will arrive at 50 trillion gigabytes. Having the option to saddle the intensity of this information through data science is incredibly significant. Data science systems can investigate historically, cause correlations with rivalry, to break down the market, and eventually, make proposals of when and where your product or service / administrations will sell best. Data scientists additionally use an algorithm like XGBoost which is an ongoing and effective development. By utilizing data, advertisers would now be able to follow clients along the way from starting enthusiasm to conclusive buy. With bits of knowledge driven by site treats and active visitor clicking percentage (CTR), advertisers have a much clearer image of what's working and what isn't, enabling them to organize use in the correct channels.

Herbert Marrow, an HR Director at Careers Booster, recently noted: "Companies that embrace data science spend less time and money on HR management – and they do this simply by using data that has always been available but no one knew how to take advantage of it." "Worldwide revenues for big data, data science and business analytics will grow from $130.1 billion in 2016 to more than $203 billion in 2020" by Banking and Manufacturing Industries (Framingham).

On February 6, Facebook propelled another call to welcome the scholarly network to submit recommendations for the Statistics for Improving Insights, Models, and Decision's demand for proposition (RFP). The RFP is a joint exertion drove by the Facebook Infrastructure. Data Science and Core Data Science (CDS) groups and expands upon the 2019 research grants right now. "We are interested in leveraging recent advancements in statistics and machine learning to improve the performance, reliability, and efficiency of Facebook's infrastructure," says Rajiv Krishnamurthy, Research Data Science Director. "With this RFP, we hope to connect with academia on recent topics and foster new collaborations."

The Face book CDS group is an innovative work group taking a shot at improving Facebook's procedures, foundation, and items that empower more than 1.5 billion individuals to speak with each other consistently. An objective for CDS is to be on the cutting edge of advancement and join forces with the scholarly community to create best in class calculations and philosophies. "I've been particularly excited by our work on measurement and evaluation under uncertainty because we can't solve what we can't measure," says Aude Hofleitner, Facebook Manager and Data Scientist. "For that reason, we're quite excited about receiving proposals in the areas of learning and evaluation under uncertainty, as well as efficient sampling and prevalence measurement. "We're also focusing on understanding how behaviour and interaction can be indicative of malicious activity, and we're looking forward to working with researchers who submit proposals in the areas of statistical models of complex social processes and anomaly detection," Hofleitner.

## UTILIZATIONS OF INFORMATION/DATA SCIENCE

- Engaging Management and Officers to Make Better Decisions
- Coordinating Actions Based on Trends—which thus help to Define Goals
- Provoking the Staff to Adopt Best Practices and Focus on Issues That Matter

- Recognizing Opportunities
- Dynamic with Quantifiable, Data-driven Evidence
- Testing These Decisions
- Distinguishing proof and Refining of Target Audiences
- Enrolling the Right Talent for the Organization

Data science is a field which uses algorithms processes and other scientific method in order to extricate information and the bits of knowledge from data in different structures both organized and unorganised. Data science consolidates measurements, information examination, AI and other related strategies so as to comprehend and break down the data for successful functioning of any business as shown in Figure 1 below (Patel, 2018) and presented applications on data analytics in table 1. The most well-known models incorporated in modern management: Hadoop, Spark, Hive, Pig, Drill, Presto, Mahout, etc. At last, data researchers should realize how to access and inquiry a significant number of the top RDBMS, NOSQL, and New SQL database the board frameworks.

Table 1. Applications of data analytics

| | |
|---|---|
| **Finance** | Constant investigation of investments |
| | Hazard or fraud investigation |
| | Shopper investigation |
| | Extortion location |
| | Algorithmic exchanging |
| | Return and risk analysis |
| **HR** | Reduction cost of awful contracts |
| | Increment degrees of consistency |
| | Foresee execution |
| | Improve benefits bundles |
| | Legitimate and moral issues |
| | Legitimate and moral issues |
| | Hierarchical development |
| **Marketing** | Client division |
| | Prescient investigation for clients' conduct |
| | Prescient investigation to carry the correct item to the market |
| | Suggestion motors |
| | Market bushel examination |
| | Improvement of advertising efforts |
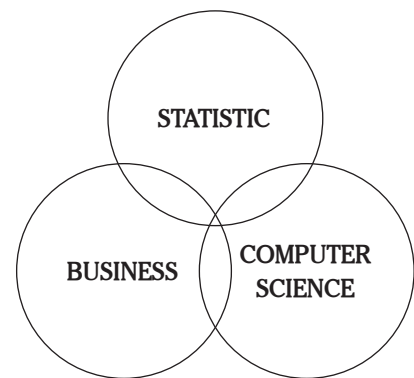


Figure 1. Data Science



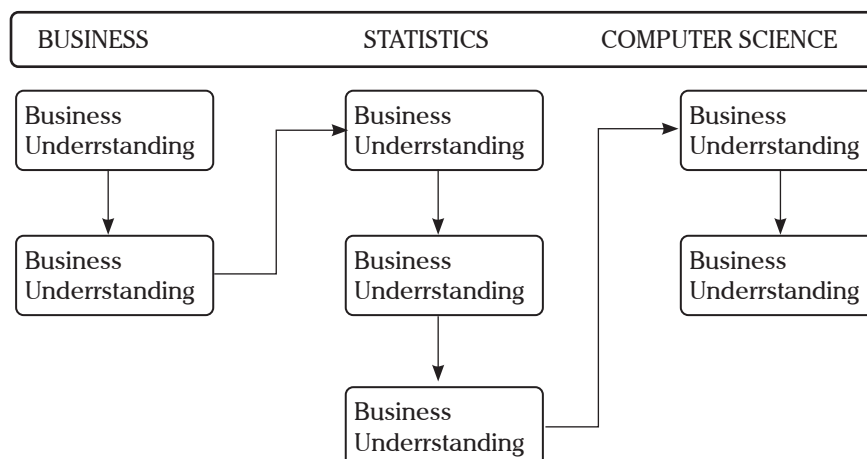Figure 2. Data Science Process

There are three steps involved in data science process which can be seen in Figure 2.

(i) Determine: The primary objective is to understand what the client wants to achieve

- To achieve customized targeting.
- To prevent fraud or unethical practices.
- To recommend products.

(ii) Understand:

- Identify the secondary or competing objectives.
- List assumptions constraints and other important factors.
- Understand the success criteria.
- Study all the possible existing solutions if any.

(iii) Map: (Business objective-technical objective)

- State the primary objective in technical terms.
- Describe how the data science project will help us solve a business problem.
- Explore all the successful scenarios.

## FUNDAMENTALS OF DATA ANALYSIS

BY 2020, each individual on the planet will make 7 MBs of information consistently. We have made a bigger amount of data in years of the whole history of mankind. Big data has surprised the world and there are no indications of easing back down. But with the help of data analysis superior leaders can make/ease out strategic decisions with the quantifiable proofs in the sectors of human resource, marketing, finance as well in the branch of productions and operations for the successful run of business. Organizations currently approach more data and analyse that data with respect to their clients than ever before. Just as surface-level information about their sexual orientation, age, and land area, organizations would now be able to infer progressively definite bits of knowledge in regards to customer conduct and inclinations.

With respect to human asset management, People/employee related information is utilized to all the more likely comprehend the association's human capital, workforce limit, hazard, and business execution. Workforce planning with the tools of data analysis now becomes an interminable procedure for deciding and surveying the necessities of an association and its workforce. The data analysis reduces the expenditure of wrong recruitments, saves lot of time, Increment degrees of consistency and also foresees employee's performance execution for a period of time.

For over five decades, advertising has been a significant field of utilization for a large number
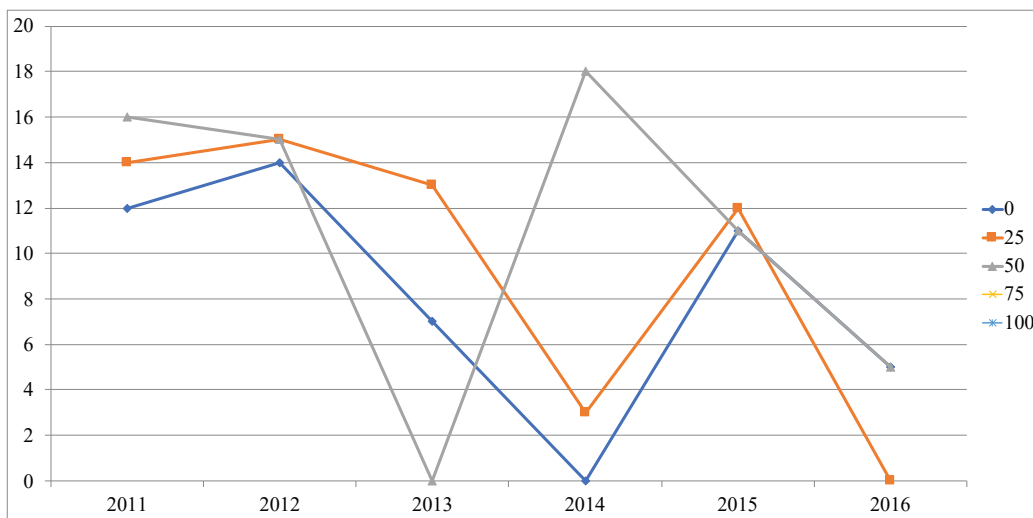


Figure 3. Expected Trend of Marketing of Data examination during 2012 and 2019

of information examination and characterization strategies. The complex assortment of satisfactory information from and about clients and contenders just as the intentional use of strategies like group, discriminate, factor and relapse analysis, multidimensional scaling and affiliation examination have become generally acknowledged guidelines in the treatment of showcasing issues.

Note that blue, orange and grey lines speak to use of data analysis in business for marketing sector especially for cost benefit analysis, break even analysis, decision making and advertising. Cantering into Customer Behaviour Analytics, the procedure of this model beginning from: gathering client information utilizing grocery store card information, cell phone application information, Geo-localisation information and it conceivable to include different wellsprings of information like climate information.

LeCompte and Schensul (1999) characterize analysis as the procedure a specialist uses to decrease information to a story and its understanding, Patton (1987) demonstrates that three things happen during investigation: information are sorted out data are decreased through rundown and arrangement,

and examples and subjects in the information are distinguished and connected.

Data analysis is the process of extracting the required data in order to perform particular operations based on user requirements. Several data analysis techniques help in performing the scrutiny in an efficient way by less time consuming with accurate results. With the help of different data analysis techniques, data scientists/researchers can play out specific analysis or tasks on datasets which are valuable to business association and also industrial application. Different data analysis techniques are used for different purposes based on user requirements; one should select appropriate data analysis technique to perform data analysis. In order to apply specific technique for the problem given, one should have the proper knowledge of data analysis techniques and their applications.

Data Analysis is described as a technique of organizing or cleaning, varying, and showing information to discover accommodating data for business fundamental administration as appeared in figure 4. The purpose behind Data Analysis is to remove supportive data from information and taking
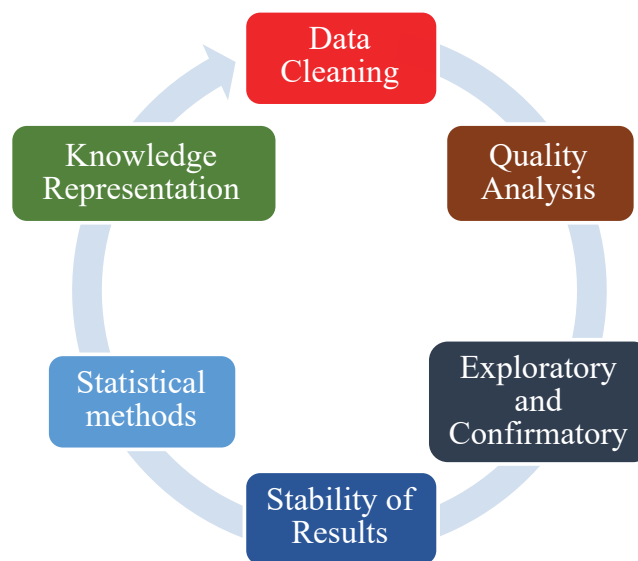
Figure 4. Data Breakdown Characteristics

the decision subject to the information investigation and assessment.

Data analysis is the act of looking into the data cleaning and converting it in order to arrive at conclusions which can help us in decision making. This is one form of analysis which is usually conducted during research experiment. Data from various sources is gathered in order to analyse the data that is acquired and arrive at a conclusion as depicted in figure 5. Data analysis can also be defined as the act of evaluating data. In today's world data analysis plays a major role in decision making which can help the business to function effectively (John, 2017).

**Steps in Data Analysis**
*Data analysis techniques*
Data analysis techniques are those techniques which help in analysing the data in an efficient way (Dana, 2019).

Mainly there are two types of analysing techniques as given in figure 6.
1) Quantitative Statistics Study (compact by quantities).
2) Qualitative Statistics Study (compact by analysis which prefers qualities over quantities).

Quantitative analysis methods are more structured compared to qualitative analysis methods.

*Quantitative Statistics Study*
Quantitative assessment is a logical and verifiable technique for thinking about direct and envisioning results that monetary experts and the board use in their essential authority process. Quantitative examination is using information from business to choose its thriving. It urges to see data to make sense of what ought to be changed in the association or what is working for the association. Specialists can use quantitative examination to choose procuring decisions, promoting or advertising decisions, and even arrangements decisions.

Quantitative data refers to data which can be counted or data in terms of numbers (ex: salary of an employee, age, amount to be paid in bank system, account balance etc). This represents "*how much*" or "*how many*". Here while gathering the data; it can be collected in terms of numbers. Analysing quantitative data is known as quantitative data analysis. In spite of the fact that there are numerous different strategies to gather quantitative information, those referenced above likelihood inspecting, interviews, poll perception, and archive audit are the most well-known and broadly utilized techniques either disconnected or for online information assortment.

The data used in quantitative process provides an information that can be assessed and noted with
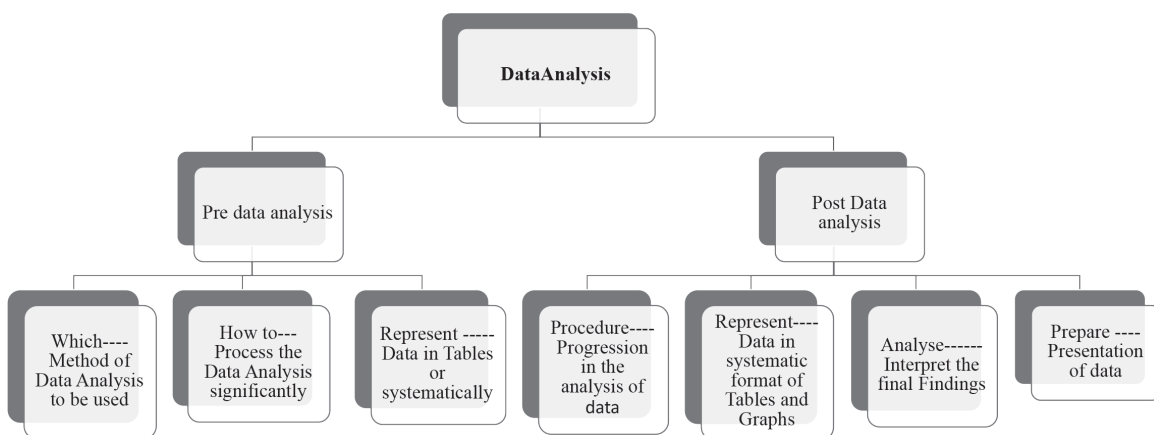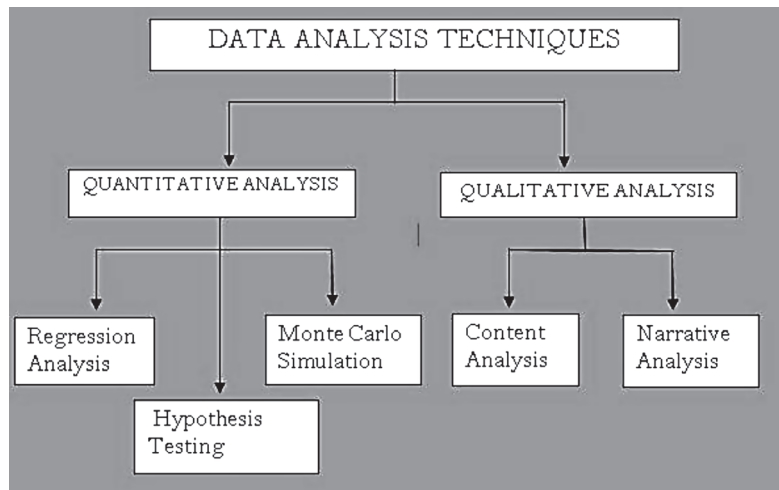


Figure 5. Data Analysis Steps

Figure 6. Data analysis techniques

Table 2: Difference between Quantitative Statistics Study and Qualitative Statistics Study

| QUANTITATIVE STATISTICS STUDY | QUALITATIVE STATISTICS STUDY |
|---|---|
| Number-Based | Content Based |
| Factual Analysis is simpler | Measurable investigation is more earnestly |
| Gathered utilizing reviews, perceptions, analyses and meetings | Gathered utilizing interviews, composed reports, perceptions |
| Information type that can be estimated and communicated numerically | Information type that comprises of distinct explanations |
| Merged (Narrow in, progressively definitive) | Dissimilar (go more extensive, increasingly explorative) |
| Deductive, centered, result direction | Inductive, all encompassing, process direction |
| A or B? what amount? | What are, who is., the manner by which do. |
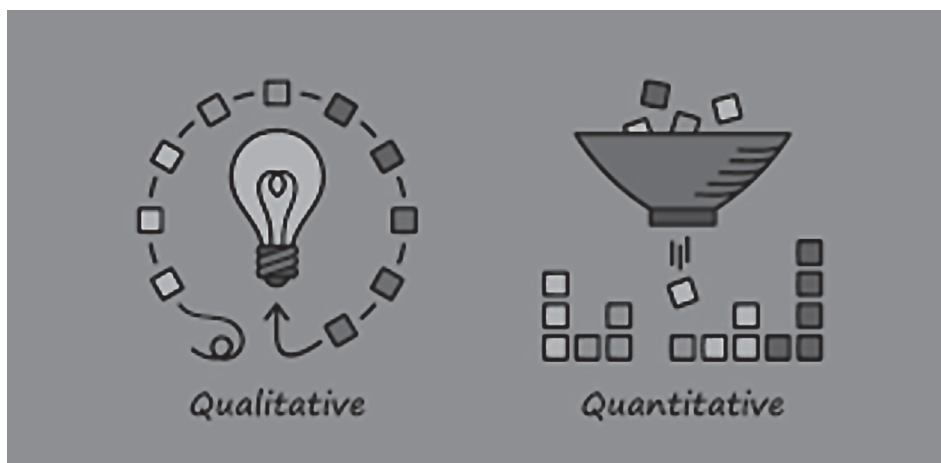


Figure 7. Pictorial representation of Qualitative and Quantitative

the numbers. A couple of existence of numerical data includes height, length, size, etc. Quantitative techniques underscore target estimations and the genuine, logical, or numerical assessment of information accumulated through studies, surveys, and audits, or by controlling earlier quantifiable information using calculation strategies. The most ordinarily utilized types of quantitative examination in business are the money saving advantage analysis, the breakeven analysis the original investment investigation, the factual investigation, and the feasibility study. Examples Organization's monetary information and showcasing returns with factual information on socioeconomics, cost of products sold (COGS), net benefits and net benefits.

### Regression analysis

Regression examination is an algebraic method to help in finding the association among two or more attributes in given dataset. In a data analysis, Regression analysis helps in determining that which attribute is more important and which attribute can be neglected. For example, consider the data set with attribute such as name, date of birth, salary, address, gender and if we want to predict a person age then date of birth matters most and attributes such as salary, address, gender can be neglected.

Regression investigation is a procedure for statistically sifting through the factors which has an effect. In factual displaying, Regression investigation is a lot of measurable procedures for assessing the data for Galton, relapse had just this organic significance, yet his work was later stretched out by Udny Yule and Karl Pearson to an increasingly broad factual. Regression is a factual strategy utilized in finance, contributing, and different controls that endeavours to decide the quality and character of the connection between one ward variable (for the most part indicated by Y) and a progression of different factors (known as free factors). Regression supports corporates and its financial directors to estimate the resources like man-power, cost, etc and realize the inter-relation between various factors, e.g., warehouse expenditure and the purchases of firms for managing in those products.

### Steps in regression analysis are as follows.

First define dependent variable (main variable based on which were practicing to forecast) with independent variable (variable that were conjecture affect subordinate variable). In order to find out whether there is a relation between dependent variable and independent variable, a graph should be plotted based on the data. Here independent variables ought to be plotted on x-pivot
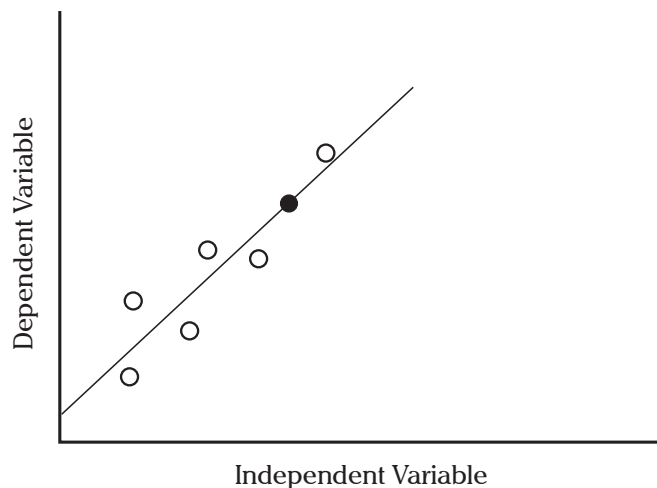


Figure 8. Regression line in graph

and dependent variables on y-hub. Once plotting is done, correlations can be observed.

Now draw regression line, i.e. a line through the middle of all points on the graph as shown in Figure 8. This line represents the relationship between dependent and independent variables and can be represented by Eq. 1.

Linear regression equation:

$$Y = bx + a \quad .......................................................... \quad (1)$$

A straight relapse line provides the structure Y = a + bx, where x is the informative variable and Y is the dependent variable. The slope of the line is b, and a, is the catch (the estimation of y when x = 0). Linear relapse measures the connection between at least one indicator factors and one result variable. For instance, straight relapse can be utilized to measure the overall effects old enough, sexual orientation, and diet (the indicator factors) on stature (the result variable). Simple linear regression: a self-contained autonomous variable is exploited to predict the valuation of a reliant variable. Multiple linear regressions: at least two free factors are utilized to anticipate the valuation of a needy variable. The contrast between the two is the number of self-directed factors.

### Hypothesis Testing

Hypothesis testing is the statistical method which is used to compare two datasets or two samples in a dataset. The use of hypothesis testing to check whether claims made on the samples are true or not. Basically, there are two opposing hypothesis tests, null hypothesis and alternative hypothesis. Whenever a claim is being made either accept it using null hypothesis or reject it using alternate hypothesis. A step-in hypothesis testing is as follows (Carolina, 2018). An null speculation is a theory that says there is no measurable criticalness between the two factors. It is normally the theory a specialist or experimenter will attempt to negate or dishonor. An alternative speculation is one that states there

is a measurably noteworthy connection between two factors.

Define two hypotheses so that only one can be right i.e., mainly defining null(H0) and alternate(H1) hypothesis as shown in Eq. (2) and Eq. (3). Null hypothesis states that the data samples which are compared is similar (no observable change).

$$H0 : \mu A = \mu B \quad ..................................................... \quad (2)$$

Alternate hypothesis states that the data samples which are compared is not similar.

$$H1 : \mu A = \mu B \quad ..................................................... \quad (3)$$

Form an analytical plan which describes how the data will be evaluated.

Execute according to plan and analyse the data sample in dataset i.e., first check whether datasets follow a normal distribution and are independent of each other. Then carry out statistical tests to verify there is difference between datasets. Analyse the result and accept or reject the null hypothesis.

### Monte Carlo simulation

Monte Carlo Simulation is mainly used to solve problems related to probability. This simulation is repeatedly use random data samples to make estimation for unknown parameters. Monte Carlo method was first discovered by Stanislaw Ulam. Monte Carlo strategies or tests are an extensive class of computational calculations that depend on rehashed irregular inspecting to get numerical outcomes. The fundamental idea is to utilize haphazardness to take care of issues that may be deterministic on a basic level.

Main advantage of this method is that many complicated problems can be solved easily. Steps with example (estimating pi) as illustrated below. Insert a circle inside square in such a way that side length of square is equal to diameter of circle as shown below in Figure 9.
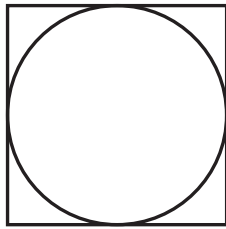
Figure 9. Circle inside square

Ratio of is of circle to are of square can be calculated as follows.

$$\frac{Area\ of\ circle}{Area\ of\ square} = \frac{\pi r^2}{(2r^2)} = \frac{\pi}{4} \quad .................. \quad (3)$$

In order to estimate the value of Pi, plot many points inside square and calculate the proportion of points inside circle to the total points as shown in Figure 10. Multiply final answer by 4 and that will be the value of Pi. As number of point's increases, accuracy of the answer will be more (Owen).
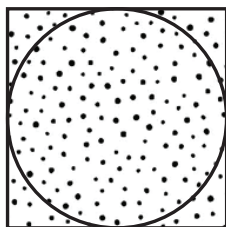


Figure 10. Circle inside square with points plotted inside square

## QUALITATIVE DATA ANALYSIS

Qualitative data refers to parts which are more interpretable and subjective. It is more subjective rather than mathematical statistical analysis. These methods are comparatively less structured than the first. Qualitative data contains data from surveys like Google forms, or employee-employer details (Dana, 2019). It focuses more on the quality aspect rather than the quantity aspect as have mentioned above. Qualitative data is characterized as approximates and describes. This data type is non-numerical in nature. This kind of data is gathered through strategies for perceptions, coordinated meeting,

leading canter gatherings and comparative techniques. Qualitative data in insights is otherwise called straight out data. One of the most important parts of Qualitative data analysis is finding patterns. So how is it done? It can find patterns first by looking at repeated words. This method is called word-based method. For example, in a talk about loan prediction we may find the words 'loan', 'interest', 'EMI" commonly.

Next is the compare and contrast method. This involves us to find in what way a text is similar or dissimilar. Therefore, it can group the data to find patterns for analysis. A lesser used method is to classify using key words used. Now can analyse by looking at the context in which it is used based on the keyword in the context. For example, when we have to analyse anger issues, we find key moments which trigger anger among the patient.

Difficulties looked in qualitative/subjective information investigation

A famous researched called Beck noted in 2003 common pitfalls that analyst create in examining qualitative data. I would like to mention a few of them. They are data shuffling, impulsive closure and overlay postponed (B, Kawulich B., 2004). Most common was related to closure when we stop collecting data because it is complex or all categories get saturated. This can be determined by trying out all assets, immersion of classes, and rise of regularities in information and over expansion. Irregularity between question and philosophy, inadequate methodological data, and absence of regard for the philosophical supporting of subjective technique are some significant difficulties (Corbin2008).

### Types of Qualitative data analysis
#### Content analysis
This method is to understand what kinds of themes are present in qualitative data analysis. It works well when we deal with data like feedbacks, interview data, and surveys from Google forms, and more. It

will be helpful to identify the most important areas to focus on for improvement. It can also use colour coding themes to find common threads in the context. In general, this method is used to analyse information obtained from research in form of surveys as text, media or sometimes physical items too. Response from interviews is analysed like this. So we can put it as the procedure for categorising verbal or behavioural data (non-numerical) for purpose of tabulation and data analysis. The process chart for content analysis is show in Figure 11.

The researcher who is directing a substance examination will utilize 'coding units' in his/her work. It varies on the kind of information utilized; however, a model would be the quantity of positive or negative reviews by an audience who are reviewing a movie. The flow chart for content analysis is as shown below in Figure 12.

Steps in content analysis are illustrated as shown below.
- Preparing data
- Deciding theme of analysis
- Categories and scheme definition
- Pre testing scheme
- Testing a sample
- Apply coding to data
- Check validity of coding
- make inferences based on results
- present results in form of graph and matrices

Advantages are that it is faster, reliable method and has ability to conduct mathematical analysis of data from the output. Disadvantages for this kind of analysis are, it only describes the data and needs to be extended as there is no explanation for data patterns.
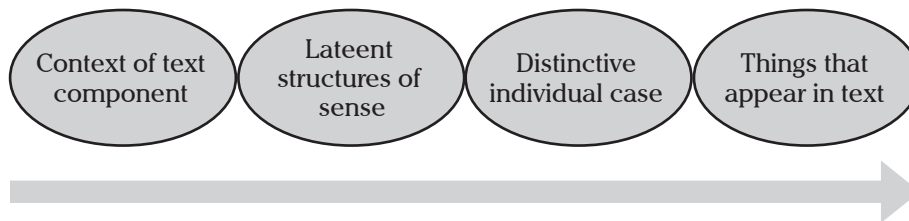
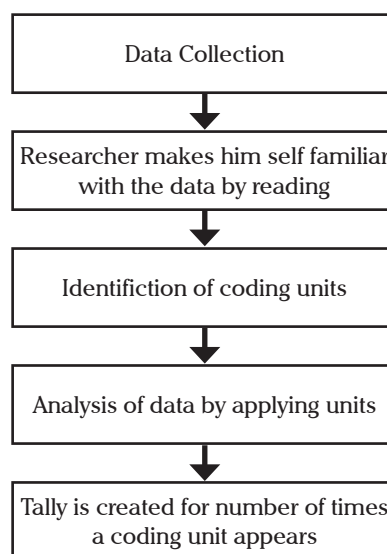Figure 11. Procedure outline for examination of data by content analysis

Figure 12. Flow chart for breakdown of data by content examination

### Narrative analysis

Narrative method is used to analyse data from interviews of employees, observations from the field, or surveys from forms and various other sources. It is based more on the real-life experience of people using their stories to help in research (Bhatia, 2018).Narrative as it sounds is dependent on narration of people in their own way and interpretation of data is done based on it and it focuses on how stories are communicated to better understand the culture of an organisation. It includes interpreting employees feeling about the work load and pressure at work to get an image about how employees and customers perceive an organisation as a whole to the public eye. It also includes how operational processes are viewed. It can be used by managers to make proper changes to corporate culture and also planning new strategies for employees and better marketing. There is no best method for data analysis. It's the way which suits the situation that makes us choose a method. The method should always reflect data that you have collected and show insights whichever are needed to be extracted. Matching right data and analysis helps to get better insights and make educated decisions to optimise organisation based on personal. Story examination is a kind of diagnostic casings whereby analysts decipher stories that are told inside the setting of research and additionally are partaken in regular daily existence (Bailey).

### MANAGERIAL IMPLICATIONS

The proposed research emphasizes on how effective data analysis is carried out both qualitatively (Content analysis & Narrative analysis) and quantitatively (Regression analysis, Hypothesis testing & Monte Carlo simulation). A thorough comparison between these two techniques is showcased with necessary pictorial representations. In today's era, it is extremely important to analyse data in the systematic way. Applying an appropriate Statistical Method is a key in data analysis. From the proposed article, a researcher can do data analysis based on the respective gathered data sets. The authors do not have any conflicts of interest to declare.

### CONCLUSION

Data analysis techniques helps analysing the data which is available and further helps in making prediction on what is going to happen taking in to consideration the history of the available data and the trends the data has shown. These data analysis techniques are mainly used in business sector and industrial sector. Technique like Monte Carlo can help in solving complex problems related to probability. Mainly in business sector, one with good knowledge of data analysis can predict the sales for upcoming years based on the data available (sales of past few years) and can determine major factor on which sales is depend on. In industrial sector data analyst can predict performance of machines for given load based on data sets available by using suitable data analysis techniques so that profit can be maximised and cost of production can be minimised. Different data analysis techniques play a major role in this modern world. ■

## REFERENCES

B, Kawulich B., Data Analysis Techniques in Qualitative Research», Journal of Research in Education, vol. 14, no. 1, pp. 96-113, 2004.

B. Carolina, Towards Data Science», Hypothesis Testing In Real Life, vol., pp. 5-9, 2018.

B. Owen, Monte Carlo Theory», International Conference on Monte Carlo and quasi-Monte Carlo methods: MCQMC 2016.

D. John, The Data Analysis Process: 5 Steps To Better Decision Making», Big Sky , 2017.

Elgendy, N.: Big Data Analytics in Support of the Decision Making Process», MSc Thesis, German University in Cairo, p. 164, 2013.

Famili A, Shen W-M, Weber R, Simoudis E. Data preprocessing and intelligent data analysis», Intel Data Anal. 1997;1(1-4):3-23.

Kelly J, Floyer D, Vellante D, Miniman S. Big data vendor revenue and market fore?cast», PP.2012-2017, Wikibon, Tech. Rep. 2014.

L. Dana, 5 techniques to take your data analysis to another level, tech talk», Sisense Publications, 2019.

Lee, R., Luo, T., Huai, Y., Wang, F., He, Y., Zhang, X.: Ysmart: Yet Another SQL-to- MapReduce Translator. In: IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 25-36, 2011.

M. Bhatia, Your Guide to Qualitative and Quantitative Data Analysis Methods», SocialCops Publishers, 2018.

T. Jolliffe, Principal Component Analysis», Springer, New York, 2002.

V. Patel, The Data Science Process», Derive Publications, 2018.

Z. Hongjun, H. Wenning, H. Dengchao and M. Yuxing, Survey of research on information security in big data», Congresso da sociedada Brasileira de Computacao, pp.1-6, 2014.