

# A Comparative Analysis on Various Business Intelligence (BI), Data Science and Data Analytics Tools

Koustubh Sharma<sup>1</sup>

Master of Business Administration in Technology  
Management, SVKM's NMIMS University, Mumbai India,  
Koustubhsharma20@gmail.com

Aditya Shetty<sup>2</sup>

Master of Business Administration in Technology  
Management, SVKM's NMIMS University, Mumbai India,  
adityashetty60@gmail.com

Arnish Jain<sup>3</sup>

Master of Business Administration in Technology  
Management, SVKM's NMIMS University, Mumbai India,  
ashutallera.ac@gmail.com

Ritesh Kumar Dhanare<sup>4</sup>

Assistant Professor, Mukesh Patel School of Technology,  
Management and engineering, SVKM's NMIMS  
University, Mumbai India,  
Ritesh.dhanare@nmims.edu

**Abstract**— Data plays a very important role in the growth of any type of business since through analyzing the data, the structure of further work and proper planning and execution can be made. Decisions based on data have a high probability of success and will help in growing the business. Business intelligence as well as data analysis helps in analyzing the data in order to gain insights on the dilution. Both Business Intelligence and data science revolve around data. While data science is the larger pool that contains more information, it can be considered as part of the larger picture. In addition, the scope of business intelligence is limited. Today, the terms "business intelligence" and "analytics" are used to some extent interchangeably. Both terms extensively describe the practice of making better business decisions through the use of large data. But are the two technologies the same? The synchronization of BI and Big Data Analytics is needed for a better decision-making process. They both are not the same thing, but have many common objectives. Today, companies around the globe are rapidly changing and companies have realized that they are not only dealing with selling records or structured data, which would otherwise be processed, used, analyzed and achieved. BI & Big Data Analytics aims to explore why, what and how, customers, products, employees and businesses are involved in sales. However, in this paper you can find out how organizations meet these expectations, what the way ahead is for BI, future challenges, opportunities and the need for analytics.

**Keywords**- Business Intelligence, Data Science Tool, Analytics, BI Challenges, BI Opportunities.

## I. INTRODUCTION

The term BI is defined in many ways, but its aim is to collect hands-on information from datasets for better decision-making and organizational growth. The [3] BI is defined as a "collection of techniques, methods, structures and processes that turn large data into assets and information that can be used

to make effective decisions." BI involves the provision and integration in an organization of business information. Data mining or storage is not new, but companies have realized over the last few years that BI is much more than corporate reporting from an IT section information house. The advances in information visualization allow for accessing information and its use to reduce costs, simplify organizational efficiencies, refine and launch new products and services. BI is still one of the most rapidly used business tools. Today, most companies have multi-regional databases and vendors of several databases with peta bytes of data spread all over the world. These databases store and manage data over the last three to four decades. Many companies have developed from these permanent data bases to complete data warehouse platforms and analytic system. Leading companies such as Wal-Mart, even before semi-structures and unconstructed Big Data analytics were presented, are now of billion dollar value in long retail chain. From this, we can say that it was not only information that led the company but also the data analysis as a service helped in attaining this target. There are various other companies which can be illustrated to validate this fact. The Amazon was a product company for online e-commerce when it began. However, today Amazon is viewed as a leading service provider for cloud infrastructure, platform and software. As one of the leading cloud service providers on the market, Amazon also offers big data. Not only does Amazon have such large companies offering a wealth of services, but Microsoft Azure, IBM Cloud, Google Cloud Platform, Sales Force and many other businesses are leading the market.

## II. EXISTING TECHNIQUES

The importance of data in the industry is very important, because only by analyzing data, a company can make substantial and workable choices, which will help it to grow

further. The data analysis is supported by data dilution information, enterprise intelligence and data analysis. The words 'business intelligence' and 'analytics' today are used to a certain extent interchangeably. In general both terms are described as practicing the best business decisions by using Big Data [1].

Business analysis and big data are trends that positively impact the worlds of business. Previous studies present the enormous and exponentially growing data generated in the modern world. They comprises of organized and unstructured data which is received by the organizations daily. The world's most digital data includes unstructured information comprising of films, pictures, emails, Multimedia messages, audios, web media post, text files, etc. The unstructured data can't be managed by the traditional Database Management System (RDBMS). Duplication of data thus calls for a review of techniques of data collection, processing and storage. Big data took on this role [3].

The study has shown the value and opportunity to create business value with BI&A technology for more information, organizational resources which sums up to current literature with: (1) empirical data on the defined resources and IT capabilities in Wade and Hulland (2004). (2) Identify and describe critical organizational resources and capabilities which worked on building and developing BNDES capabilities with the former. This study therefore supports the Melville, Kraemer and Gurbaxani models (in 2004) and the New and Wade models (in 2010) and validates the emperical findings of Nevo and Wade which was presented in 2011 [4].

A large number of data, such as text, audio, video, picture, etc., are currently being generated in different forms. With this large volume of data, it is very hard to handle and look for correlations, patterns and huge data sets. This article provides information on analytics of big data. Different tools and methods are discussed for big data analysis. Hadoop can be a better option for workloads in batches only making it probably less expensive in comparison to other solutions. [5]

Homogenous clusters were found by the latent clustering technique. Measures have been modeled in poLCA analysis for RStudio 1.0.153. The two-block models were selected since it had the lowest measurement of the Bayesian BIC and the highest probability. Algorithms and analytical skills are the best differences among the clusters. The differences among each complex are therefore strong. Big data analytics Cluster 1 was named and business intelligence Cluster 2 was designated [6].

This article describes in detail the methods of big data and large data analyses, the technologies used by various companies and the media with their properties used in the market. Different events and data from industry professionals have thoroughly analyzed the challenges and future trends for big data. The key to this article is summarized: Big data is the biggest tool for companies in the current scenario to capture and discuss systems. Access is provided for every type of data from different locations around the world with the technology and tools which are in use at present. Companies can benefit from data gathered with feedback readily available, automated

processes, for getting information on their target market and for improving overall performance [7]

In contrast, the rate at which information management tools are developed is significantly lower than that of information development. In the current market, not all the issues of big data analysis are addressed. Indeed, even advanced tools and techniques like Hadoop, Cassandra and Ignite cannot justify an analysis in real time. It is important to increase the stress-free handling of different data sets and in addition, to reduce the time taken for processing. Some storage, search, analysis, sharing and security problems have still not been resolved. In the future we will improve and develop tools for large-scale data analysis [8].

The researchers concluded that features, quality and visual attraction increase the open source business intelligence platforms. In order to analyze crime data from the six testing tools, this report recommended that Apache Hadoop include not only distributed storage and management of large data sets but also the ability to store and process large volumes of data in a rapid way. Apache Hadoop has been recommended by this study which provided quick computing power, tolerance of failure, flexibility (procedures) [10].

Popular data analysis instruments and their functions are presented in this article. The advantages and limitations of each body are discussed. This article clearly demonstrates the best suited tool to be used for analytical work. The tool can be easily verified by researchers/data analysts from discussions to meet their needs [11].

Business analysis is seen as a great advantage for enterprises, as it contributes for informing competitions on time, optimizing business processes and creating opportunities for growth and innovation. When business analytics initiatives are implemented by companies, there are numerous strategic issues, such as how business analytics maximizes the value. The role of corporate analytics and the need for corporate analysis have been the focus of recent information systems' (IS) literature. Theoretical and practical success factors related to the operation of business analysis have, however, received little attention. The principal aim of this study will be the empirical analysis of the success factors of enterprise analysis and the impact of this gap in IS literature presented over the enterprise analysis organizations. Through a qualitative study, we have gained an overview of the results of the analysis and the success factors. Our research informs and helps to shape business analysis theoretically and practically [12].

The data have been dramatically generated in recent years. For a general person it is hard to analyze these data. In this paper, the various issues, challenges and tools that have been used for analysis of Big Data were examined. This research understands that every large data platform has its own focus. This research is designed for bulk treatment where others can be analyzed in real time. Each large data platform also has specific functions [13].

The Internet of Things (IOT) miniature has increased in the last decade, generating large volumes of data. But without analytical power, such data are not useful. Value full information about extensive data generated by IOT devices can

be obtained from a lot of large-format data and IOT solutions. However, these solutions are still early, and there is no comprehensive field research available for the same. This article explores recent efforts in technological research in large-scale IOT data analysis. This is why IOT analysis and Big Data are related. Furthermore, this article adds value to propose a new architecture for broad IOT data analysis [14].

Author examined Hay's innovative subject Big Data in this research. Recently, the authors were very interested because of its unforeseen advantages and possibilities. A large quantity of high speed data, including details and hidden patterns of knowledge, is generated every day in this information age. This enables Big Data Analytics to use advanced big data analytical techniques and reveal hidden information and valuable knowledge, to improve the decision making process [15].

To match the high data rate, various aspects of high-speed data is searched. While the research gives a sense of advancement in big data research, some research issues in this field remain open for future research. Problems like interoperability of data and tools for various data formats and the integration of various analytical frameworks for large data are increasingly addressed. Security and privacy are another major concerns for researchers and leads the retention of sensitive information that was communicated, particularly in the cloud. In the future, streaming data downloads will require effective frameworks for real-time analysis. A challenge requiring a lot of research is effective safety monitoring [16].

A summary of modern technologies and data analysis trends is provided. This document provides a summary. It collects stores, merges and sort enormous quantities of information. Software and hardware systems were a major challenge. The increasing number of companies has developed and solved tools for storing tables, documents and multimedia data. In common applications, database structures are a key tool. Thousands or millions of entries in these structures are available every day. The purpose of analytical tools is to obtain information needed and is useful for active control and decision-making. This contribution aims primarily to present a range of data analysis capacities and tools regarding the availability of users [17].

Business intelligence has two main features for companies. On one hand, it reduces costs and on the other, increases revenue. BI tools can support management strategies and options significantly. The BI provides an analysis of data to support products, service and/or marketing and related processes in their development, improvement or redefinition. This allows any organization or company to benefit from the BI application irrespective of size, business or placement on domestic and international markets. The choice of the most appropriate BI platform for a company is based mainly on the CEO and/or company entrepreneur's visions, such as current financial condition, human or material resources. [18].

Big data arrival in society is because of the value and use of business and government action. This article described Big Data characteristics and presented the big data analysis architecture. Large data technology differs from traditional SQL-based RDBMS approaches for broad volumes, speeds and variety of data. The new standard includes NoSQL database,

huge parallel and scalable computer platforms, and open source software [19].

Lastly, after analyzing big data instruments in open and closed sources, the use and needs of an individual or company become clear. There are a few personal instruments that cannot be offered due to prices and complications, but open source systems may pose a problem of obsolescence and modification [20].

### III. TYPES OF ANALYTICS

The use of affiliations is combined with three types of analytics: clear, prediction and precautionary. Clear analytics use corporate intelligence and mining records to report trend facts on past or found activities. Detailed analytical training is carried out right down into data to unveil information such as frequencies, function prices and the root cause for analysis (IBM, 2013). Conclusive analytics provide remarkable understanding of business efficiency and allow customers to keep an eye on their business activities as well. The preparation of analysis uses a variety of styles and additional approaches, based on historical and current reports, in order to anticipate future results (Gandomi and Haidar 2014). Data is collected in anticipation, an analytical type is made, prophecies are created, and a model is sanctioned, as files appear to be included (Gartner IT Reference. n.d). Big data are transformed into meaningful, useful business information (Abbott, 2014). It distributes the electric energy of data and also allows selected workers to take advantage of information about how to foresee future individual routines (Lustig and so on, 2010). Prescribed analytics allow decision-makers not only to look at and discover the potential for future important operations but to use them in a proper manner and most importantly on time. But in addition, it is also used to provide the absolute most efficient solutions.

#### A. Machine Learning

Different software which can learn from the data can be included in the Machine. It provides computers with the ability to learn without being programmed explicitly and concentrates on predicting the properties which are known. Presently, the use of machine learning is made for finding differences among non- spam and spam messages. This is done for learning the preferences of the users and for making significant suggestions of the basis of the attained information, etc.

#### B. Analytics Techniques

A framework that can be considered as a cycle with various stages is required to provide clear insights into customers with the help of the data. This framework includes several measures to analyze the data. Figure 1 illustrates various data analysis life cycle phases and the intermediary data flow, identify the problem, prepare data, plan and build model, communicate the results obtained with a product operationalization.

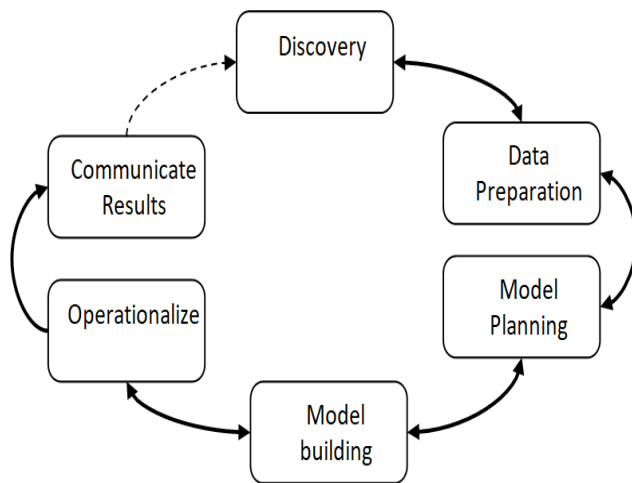


Figure 1. Data Analytics Life Cycle

- i) **Discovery** - It is very important to understand the problem and to determine if sufficiently available information allows an analytical plan to be prepared and shared for another study. The company wants to predict the data for decision-making at this stage. Analysis is therefore conducted via the datasets. That is why, the team evaluates people, technology, time and data. The issue and early formulation of hypotheses are also included in this phase.
- ii) **Data preparation** - It is important that the available data be verified whether the model is good quality or not. Data are included in the steps for exploring, pre-processing as well as arranging data. Execution is required by the ELT phase, extracting, loading and transforming. At this point, the team should know the data in well-ordered format or should know it systemically.
- iii) **Model planning** - It is crucial to determine if there is a good idea to try a model. At this point, the team identifies the process, methods and order for building up the model which is to be followed. The group identifies the information to be provided in relation to the variables, the way they get linked, and key variables are selected, thus proposing appropriate model.
- iv) **Model building** - The planned design is robust or cannot be monitored. Data sets to train, produce and test at this stage have been developed by team members. The constructed model is implemented at this stage in the planning stage to aid. The team also examines the adequacy of current tools to implement the proposed models.
- v) **Operationalize** - at the time of this phase, technical documents, project codes, summary and final documentations shall be provided by the team.
- vi) **Communicate results** - The results of an analytical plan at the discovery stage fail in this section of the life cycle.

#### IV. DATA ANALYTICS & OPEN SOURCE VERSUS COMMERCIAL SOFTWARE TOOLS

Several organizations are trying for cost cutting in their major Business Intelligence (BI) projects and are hoping for more cash from the open source BI. The main advantage which can be gained from the open source tools is to be free and also to permit the source code to be accessed by modifying the various modules. Open source platforms have the advantage of being able to replace them at no cost if they do not satisfy an organization's needs. In comparison to the commercial, generally open source tool needs lower systems, on the assumption that people cannot invest in hardware until they make investments in software. This is justified. Open Source BI platforms are assumed of evolving a lot speedily in comparison to the business platforms as they aren't restricted by rigid architecture and compatibility issues. Open source tools usually have less system needs in comparison to the commercial usage, which are acceptable if people who can't make investments in software cannot make investments in hardware. It's anticipated that open-source BI will develop a lot speedily in comparison to the commercial platforms as compatibility issues and rigid (or even outdated) architectures are not restricted. Data analysis is a process of investigation of data sets for drawing conclusion for the information they contain through specialist systems and software. Incorporated business decisions and scientific models, theories and assumptions made by scientific and researchers have commonly been used by organizations in commercial industries to verify or refute database analysis technology and technology. The data analysis can support companies in increasing their income, improving operational efficiency, optimizing marketing campaigns and efforts for customer service, responding to emerging market trends faster and increasing their competitiveness over all to enhance business performance. This can improve the operation of companies. Depending on the concerned applications, the data analyzed may consist of historical records or new data processed for real-time analysis purposes. It can also be a combination of internal and external sources of information. Crime analyses are used to analyze, map and view crime incidents or criminal patterns to provide an insight into crime prediction. The analysis of criminal information therefore helps the security and the police to take account of their resources in preventing crime.

#### V. OVERVIEW OF THE TOOLS

Good software should be used to analyzing the data and for extracting related information from business. A large number of tools are now available for open-source data and commercial analytics. To find new opportunities, you should invest in appropriate tools and skills. Software data analysis tools utilize different methods to manipulate data sets and to find suggestive results. In some ways, summary reports and improved viewing are even done well, which helps to achieve better results.

##### A. POWER BI

Power BI is a company analytics service from Microsoft. It seeks to create end users with an easy-to-use interface to

interactively view and business intelligence capacities, and its own reports and dashboards. In addition to the desktop interface, Power BI provides cloud-based BI services termed as Power BI service. Interactive dashboards, data discoveries and capabilities for preparing data are provided by the data warehouses [2]. In March 2016, Microsoft launched additional services in March of the year 2016, under the name “Power BI” Embedded on its Azure Cloud platforms [3]. The capability of loading customized visuals is one of the product’s major differentiator.

#### 1) Power BI Desktop

Mainly used for PCs and desktops, Windows desktop application for reporting on services.

#### 2) Power BI Service

It is SaaS-based online service (service software). It was known as Office 365 Power BI or just Power BI.

#### 3) Power BI Mobile Apps

Windows, iOS and Android phones and tablets Power BI mobile apps

#### 4) Power BI Gateway

Portals to sync and automatic refresh external information from and into Power BI. In enterprise mode too, flows and PowerApps can be used in Office 365.

#### 5) Power BI Embedded

The Power BI Rest API allows users of Power BI and non-Power BI to create dashboards and reports

#### 6) Power BI Report Server

A Power BI Reporting solution in an on-premise for companies that cannot or will not store data in the Power BI service

#### 7) Power BI Premium

Capacity based offering, which enables flexible reports throughout an enterprise, without requiring the individual user authorization of recipients. Scale and performance are higher than shared BI.

#### 8) Power BI Visuals Marketplace

On the market there are custom visuals and visuals powered by R.

### B. GOOGLE ANALYTICS

Google Analytics is an incredible source of information on traffic, but the correct Google Analytics analysis can tell you more than how much traffic your website transmits.

There are all kinds of information on which websites have a connection, such as traffic sources, which social network channels directly bring your visitors to the website, which keywords have been ranked on Google and therefore traffic, etc.

In addition to traffic shortages, Google Analytics interpretations can also demonstrate that visitors to your website report on key areas such as:

1. The most popular pages on your site is your top content,

2. New visitors over certain times give you a sense of how stubborn your site is and how many visitors are returning to visit us in future.
3. The time-duration which is spent by them on your site,
4. And a great deal more.

### C. THOUGHTSPOT

The Company introduced ThoughtSpot Monitor, an information monitoring tool for changing patterns or trends, in its ThoughtSpot 6 software in 2019. ThoughtSpot offers the ability of non-technical individuals to conduct their own data analysis searches. It also allows you to view and analyse your data via a search interface. You can quickly and easily create your search by simply typing it into a search bar as you do with an Internet search engine. ThoughtSpot facilitates the viewing of your data, the reply of your questions, and the creation of interactive graphs. Collaboration and security features facilitate the protection and safe sharing of sensitive data with others.

ThoughtSpot allows managers to modify data properties to satisfy business needs. Administrators can specify common terms for the search synonyms; increase the importance or the format of data in a column for search results. Therefore, you will check with your ThoughtSpot administrative to see if some settings can be changed when you don't receive the answer you expect when you use ThoughtSpot.

#### 1) Finding your way around

To make navigation easy, we organized ThoughtSpot into several sections. You can see them on the menu bar.

#### 2) About the user profile

The user icon lets you view your profile, or sign out of ThoughtSpot.

#### 3) Understanding privileges

Your privileges determine the things you can do. ThoughtSpot sets privileges at the group level.

#### 4) About stickers

You can create stickers to make it easier for people to find data sources and pin boards.

### D. QLIKVIEW

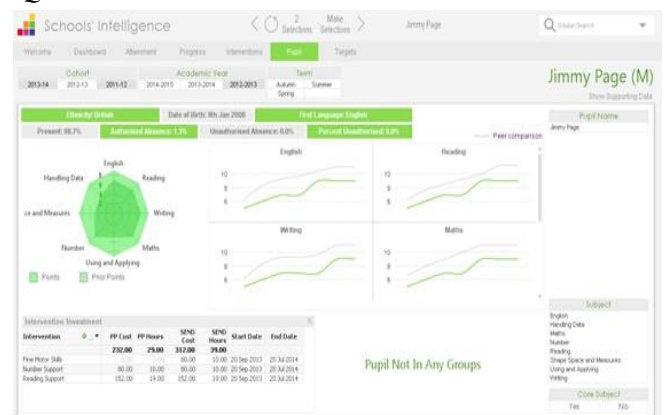




Figure 2. User Interface of Qlikview

QlikView is an important platform for business discovery. In comparison with the traditional BI platforms, it is unique in many ways. As a tool for data analysis, the relationship between data is always maintained and colours can be seen. The data that are not related also appears. It offers direct as well as indirect searches by searching individually in list boxes. The central and patented technology of QlikView has the in-memory data processing feature that gives users superfast results. It calculates fly aggregates and compresses the figures to 10% of the original size. None of QlikView application's users and developers manages the data relationship. It is automatically administered.

### 1) Features of QlikView

- QlikView features patented technology that allows you to quickly create advanced reports from multiple data sources with a variety of features. The list of features below is very unique to QlikView.
- Data Association is automatically maintained – QlikView recognizes the relationship between each piece of data in a data set automatically. The relationship between different data entities must not be preconfigured by users.
- To provide ultra-fast experiences to the users, data is kept on the memory of several users: The structure, data and report computations are all stored in the server's memory (RAM).
- Aggregations on fly are calculated as required: the fly is computed on fly when the data are kept in memory. No pre-calculated aggregate values need to be stored.
- QlikView uses strong data dictionary to compress data by up to ten per cent of original size. For any analysis, only essential pieces of memory data are required. The original data is therefore compressed to a very small extent.
- The relation between data is not displayed by arrows or lines but by colours. Selecting a data piece gives the associated data specific colours and unrelated data another colour.
- Direct searches and indirect searches – Installation of some data and the exact results can occur due to the data association rather than giving the direct value a user is looking for. Naturally, you can also directly find a value.

### E. R DATA SCIENCE

R is a software program developed by volunteers for the science, researchers and data analyst community and maintains the statistical computing foundation of R. The GNU General Public License for R is freely available. It is widely used by statistics and many guidelines are available online. Interested readers can consult R-bloggers, uc-analytics, introductory sites, and r4stats.com websites for additional research. The KDnuggets survey showed that R is better used between tools for data analysis. It says that 49% of the voters use R and that there was a 4.5% increase between 2015 and 2016. The study showed that R is being used more and more. In recent years, the popularity of R has increased significantly, according to

surveys, surveys, and database literature studies. For 40 years, the development and use of R has been challenged by SAS language. R is developed using C, Fortran and R. The software environment. R has a good interface for the command line, and it is easy to implement and understand all R commands. A number of graphic front ends are now available, such as the R studio, IntelliJ, Visual Studio, etc.

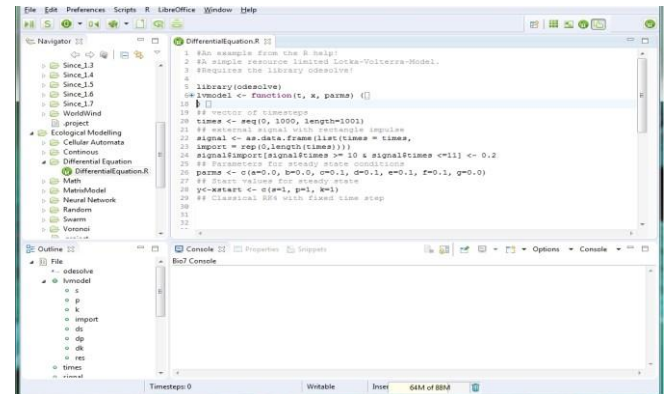


Figure 3. R Studio IDE

Figure 3 exhibits the popular IDE, R Studio. The overview of R Studio IDE which integrates variables, results, workspace, etc. in Figure 3. R is a language that has been translated. Users generally use an interpreter for a command line to access it. The RMD "R Mark Down" is short for an easy exchange of code with the output. The export of RMD files such as pdf, word, HTML and latex file is possible using knitr. There are nearly 12,000 statistical, graphic and analytical packages[23]. R can be enhanced easily by functions and the R community is known for its package contributions. Certain libraries that have shown themselves to be useful in analysing them are dplyr, tidyr, rgl, stringr, graph, ggvis, RMySQL, rMarkdown, Ggplot2, etc.

The R markup can be used to create your data through changing your analyses into dashboards, reports, presentations and documents to tell a story. R strengths include static graphs to generate quality graphs for publication. Extra packages provide dynamic and interactive graphics. The tool has few more strengths, promotes reproducible research and records how analysis has been carried out. You can modify re-run, forward, comment etc. Commands: More software runs by users, so that everyone can add new packages and enhance the code. R's restriction is that the data that can be processed is the same as the system memory if parallel processing is not used. It is not easy for the novice to use, cannot properly scale Big Data sets, and some procedures could take days.

### 1) Data Manipulation in R with dplyr

Data manipulation and clean-up are most of our time and effort in the journey from data to insights. If you use R to help you analyze your data, the dplyr packet is a life-saving device. Dplyr has been developed to allow efficient data management with speed and coding ease.

dplyr is a grammar of data handling and offers a consistent set of verbs that help you to resolve the most frequent challenges of data handling:

mutate() adds new variables that are functions of existing variables

select() picks variables based on their names.

filter() picks cases based on their values.

summarise() reduces multiple values down to a single summary.

arrange() changes the ordering of the rows.

These all combine naturally with group\_by() which allows you to perform any operation “by group”.

## 2) Tidy

Tidyr is a perfect add-on for dplyr. It increases the capacity of dplyr to manipulate and pre-process data. The following is a list of tidy functions:

gather(): The function “gathers” multiple columns from your dataset and converts them into key-value pairs

spread(): This takes two columns and “spreads” them into multiple columns

separate(): As the name suggests, this function helps in separating or splitting a single column into numerous columns

unite(): Works completely opposite to the separate() function. It helps in combining two or more columns into one

## 3) ggplot2

Ggplot2 is the best database in R to view data. In the library ggplot2 a graphics graphics is implemented (Wilkinson, 2005). This approach gives us a coherent way of visualising the attributes of the data with their graphical representation. Ggplot2 offers a variety of features.

The best visualization package I've ever used is far and away. Under ggplot2 data scientists love to produce their charts and views. The package is so useful and popular that it has been integrated into the language of Python! This package is so much for us to do. Regardless of whether it involves construction of box plots, you name it, density plots, violence, ggplot2, tile plots and time series plots has a function for it.

## 4) Shiny

The computer power of R is combined with contemporary web interactivity. Best thing-shiny applications can easily be written and developed because special web development capabilities are not necessary.

In order to increase cooperation and transparency, Shiny allows you to work and communicate on the same platform with your team. It is the ideal tool directly from R to create interactive web applications. You may host stand-alone website applications or insert them into R markdown documents. You can not only create interactive dashboards with Shiny. It includes a number of built-in input controls. Once your Shiny apps are developed they can use HTML widgets, CSS themes, and JavaScript actions.

## 5) Plotly

This is one of Plotly 's finest data visualization tools, built on the D3.js, HTML and CSS display library. The Django Framework is created using Python. You can either choose to create online interactive data visualizations or use the libraries that are offered in the language / tool of choice to create these visualizations. It can be used for several languages: R, Python, MATLAB, Perl, Julia, Arduino.

### a) Advantages and Disadvantages of Plotly

#### Advantages

- Support the various languages, such as R , Python, MATLAB and Perl, Julia and Arduino. • It is compatible.
- Interactive tracks can be shared with many people easily online with plotly.
- Plotly can also be used to create interactive plots by using data up and a plot GUI by people with no technical background.
- Ggplots for R and Python are compatible.
- It allows interactive tracks to be embedded with iframes and html in projects or websites.
- It is also very simple to create interactive plots using a plot.

#### Disadvantages

- The plots produced by plot community are always public and can be viewed by anybody.
- There is an upper limit to daily API calls for the plotted community version.
- In the Community version, there are also a limited number of colour palettes that act as a top bound to the colouring options.

## F. Python Data Science

Python is the main language developed and published in 1991 by Guido van Rossum. The language of programming is friendly, learning, powerful and quick. The OSI approved Open Source License allows this language to be used free even for business users. Python's Python Software Foundation License Management is administered. Simply PyPI, Python Package Index, hosts several third-party Python modules. The standard Python library and the modules supported by its community offer endless possibilities. Python uses 45.8%, and the survey shows a slight decrease. A large number of libraries use Python to perform data analyses. Python is very attracted to data analysts. Interested readers can refer to the data camp, the data science centre, the online byteacademy. Spyder, PyCharm, Rodeo, Atom and Jupyter Notebook are some of the most powerful data analysis python IDEs in the world. Figure 4 shows that Jupyter's notebook is not only an interactive and user-friendly environment, but also an educational tool. It is possible to share the code and its output in the same notebook as PDF and in many other files including latex.

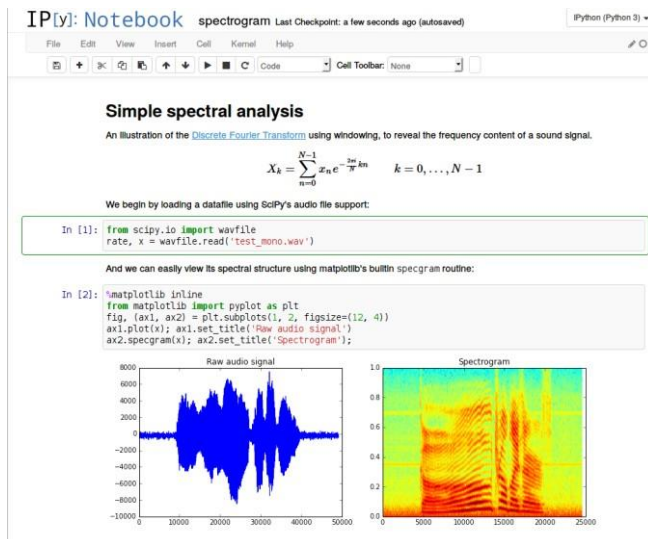


Figure 4. Jupyter notebook for python

Then libraries for data analytics include matplotlib, SciPy, NumPy, Mlpy, knowledge-learn, statsmodel, Pandas, NLTK. Pandas library is written in python and is brought in use to manipulate and analyze data. For the manipulation of time series as well as numerical data, various processes and data structures are offered by the Python. The Statsmodels module supports users in data exploration and statistical models evaluation and statistical testing. There have also been a range of scientific features in the classification, regression and grouping-algorithms consisting of vector support machines (SVM), k-means and dibscans. It can also interact with Bayes, a random forest. Mlpy is built on NumPy and SciPy that provides monitored and uncontrolled methods for machine learning. The Numpy library offers methods for supporting matrices for numerical routines that are running quickly. For integration, optimization, image processing, linear algebra, FFT, signal and SciPy supplies modules. Matplotlib is a wonderful library of NumPy plots. The most important target is of enabling the user for generating from the GUL with the tool-kits. The "Natural Language Toolkit" (NLTK) is a group of excellent libraries with sample data and graphic demonstrations. Theano is a beneficial Python library which allows us to play efficiently with the mathematical expressions, such as the definition, optimization and evaluation of arrays of all dimensions. Python's limitation as a data analysis tool is simpler to program, while R is statistically friendly. Limits to Python include lack of business support, the lack of proper support for multiprocessors and the absence of a UI development framework.

### 1) Numpy

One of the most widely used Python libraries is NumPy or Numerical Python. Having multidimensional arrays support for mathematical and logical operations is the main feature of this library. NumPy features can be used as a multi-dimensional array of real numbers to index, sort, reshape and transmit pictures and sound waves.

**A list of NumPy features is available here:**

- Conduct scientific and mathematical calculations simple to difficult
- Strong multidimensional array objects support and a number of array element functional and processing methods
- Transformations and routines in Fourier data management
- Performs calculations of linear algebra required for algorithms like Linear Regression, logist, Naive Bay, etc. For the machine learning.

### 2) Scipy

The SciPy library is a group of subpackages which contribute to the solution of the main problems related to statistical analysis. For the management of NumPy array elements, the SciPy library is used to compare the mathematical equations not possible with NumPy.

**Here's a list of features of SciPy:**

- The platform provides numerous mathematical processes such as numerical integration and optimization alongside NumPy arrays.
- There are a number of subsets that can be used to quantify vectors, process Fourier, integrate, interpolate etc.
- Provides a complete stack of linear algebra functions used for further calculations like clustering using k-means, etc.
- Supports signal processing and the creation of sparselous matrices, data structures, and numerical algorithms etc.

### 3) Pandas

"Pandas" is a large library mainly used in various fields such as statistics, finance, economics, analyses of data etc. The library uses the NumPy array to process pandas of data objects. NumPy, Pandas and SciPy are very dependent on each other for scientific calculation and manipulation.

**The following is a list of Pandas features:**

- Create quick, efficient, predefined and customized data frame objects.
- It can be used in large data sets to manipulate, subset, trim and index data.
- Provides features built in Excel charts and complete data analytics tasks such as statistical descriptive analysis, data dispute, transformation, handling, display, etc.
- Supports time series data manipulation.

### 4) Matplotlib

Matplotlib is the largest data viewing package in Python. It supports many graphs such as a histogram, bar charts, spectrum of power, error charts, etc. It is a graphic library with two dimensions that produces clear, concise diagrams, which are critical for the EDA analysis.

**A list of Matplotlib functions is available here:**

- Drawing graphs, featuring proper line styles, font styles, axles etc, is extremely simple for Matplotlib.



- The graphs created can help you understand trends and patterns and correlate them. They are usually instruments for reasoning quantitative information.
- It's very much like the user interface of MATLAB. Pyplot is equipped. It is one of the best features of Matplotlib.
- Provides a module for object-oriented graphics integration with GUI tools like Tkinter, wxPython, Qt etc.
- Provides you with a series of standard data sets. Scikit Learner Library, for example, contains renowned data sets for Iris and the prices of the Boston House.
- Incorporated methods for machine learning both supervised and uncontrolled. This includes resolution, classification, regression and anomaly detection.
- The feature extraction and functional selection features are included to help identify the important data attributes.

#### 5) Bokeh

For descriptive web browser graphic imaging, Bokeh can be used for most interactive libraries in Python. It can process modest datasets easily and build versatile charts that contribute to a high level of EDA performance. Bokeh provides the most defined interactive plots and dashboards functionality.

#### Here's a list of features of Bokeh:

- Help fast creation of complex statistical diagrams with simple commands
- HTML, laptop and server outputs are supported. Exit. It also supports several linguistic elements like R , Python, lua, Julia etc.
- Bokeh is also integrated with Flask and django, so you can also view these applications
- It enables visualisation transformation in other libraries like Matplotlib, Seaborn, Ggplot, etc.
- These were the Python libraries that were the most useful for data viewing. Let 's talk now about Python 's top libraries throughout the process.

#### 6) Scikit Learn

Python's Free Study Machine Library is Scikit-learn. It has various algorithms such as support for the vector system, random forests and neighborhoods, and also supports numerical libraries such as NumPy and SciPy by Python.

#### Here's a list of features of Scikit-learn:

- It includes cross validation techniques for the performance assessment of the model, as well as parameter tuning to improve the performance of the model.

#### 7) Keras

Keras is an open source neural network Python library. It works with Tensor Flow, Microsoft, R, Theano or PlaidML cognitive toolkit. The objective of this program was to allow rapid and user-friendly, modular and expandable experiments with deeper neural networks. It was developed as part of Project ONEIROS 'research work and its principal author and maintainer, François Chollet, is a Google engineer. Xception is also the author of Cholet's profound neural network model.

#### Here are some key features of Keras:

- Encourages the construction, interconnection, convolution, pools, recurring, embedding and so on of all types of networks; These models can be combined to form a comprehensive neural system for large data sets and problems
- It has incorporated functions for neural network calculations like layer definition, objectives, activation functions, optimizing devices and several tools for simplifying the processing of images and text data. It contains several previously processed and trained data sets like MNIST, VGG, Inception, Squeezenet, ResNet and many other.

Added new modules that contain functions and methods are easy to extend and support.

TABLE I. COMPARISON BETWEEN BUSINESS INTELLIGENCE, DATA ANALYTICS AND DATA SCIENCE

Factors	Business Intelligence	Data Analytics	Data Science
Concept	Deals with data analysis on business platform	It is used for analyzing raw data in order to make conclusions about that information	Consist of several data operations in various domains
Scope	It analyzes past data	It makes accurate predictions on the basis of events and forecasts future	It analyzes past data for future predictions
Data	Handling static and structured data	Structured but mostly unstructured data is handled	Both structured and unstructured data is handled
Data Storage	Data is stored mostly in the data-warehouse	No storage available in the tool	Data utilized is distributed in real time cluster
Procedure	It helps companies to solve questions	It answers specific questions and discover new insights for competitive advantage	Question here are both curated and solved by data scientists
Tools	MS Excel, SAS BI, Sisense, microstrategy	R Programming, Tableau Public, QlikView	Python, R, Hadoop/Spark, SAS, TensorFlow

The above mentioned table 1 exhibits similarities and differences among business intelligence, data analytic and data science. This comparison table is based over various parameters.

The Below mentioned table 2 exhibits similarities and differences among R and Python. This comparison table is based over various parameters inclusive of advantages and disadvantages of both.

TABLE II. COMPARISON AMONG PYTHON AND R

FACTORS	PYTHON	R
Primary users	Developers and Programmers	Scholar and R&D
Learning curve	Linear and smooth	Hard at the beginning
Important packages and library	scipy, caret, pandas, TensorFlow, scikit-learn	tidyverse, zoo, caret, ggplot2
Ide	Notebook, lpython, Spyder	Rstudio
Objective	Deployment and production	Data analysis and statistics
Flexibility	Easy for constructing novel models from scratch. That is, matrix computation and optimization	Easy for using available library
Integration	Well-integrated with app	Run locally
Task	Good to deploy algorithm	Easy for getting primary results
Database size	Handle huge size	Handle huge size
Benefits	<ul style="list-style-type: none"> <li>• Jupyter notebook: Notebooks help to share data with colleagues</li> <li>• Function in Python</li> <li>• Deployment</li> <li>• Speed</li> <li>• Mathematical computation</li> <li>• Code Readability</li> </ul>	<ul style="list-style-type: none"> <li>• Graphs are made to talk. R makes it beautiful</li> <li>• GitHub interface</li> <li>• RMarkdown</li> <li>• Shiny</li> <li>• Large catalog for data analysis</li> </ul>
Drawbacks	Not as many libraries as R	Slow High Learning curve Dependencies between library

## VI. CONCLUSION

Data Science can analyze past data (trends or patterns) and make future predictions. BI can help interpret past data. Power BI, Google Analytics, Thoughtspot, Qlikview, R. Comparative tools include methods and tools to analyze wide sets of sources of data for the decision-making process of teaching and learning that drives sustainable, efficient information. Based on its characteristics, specifications, advantages and limitations, the comparative analysis of several open source tools was performed. By review of various research articles found that, because of its many features, Power BI and Google Analytics is the leading open-source tool for scientists. Comparative study in this paper provides a clear concept of all the tools making it easy to various related uses.

## VII. LIMITATION

Following limitations were faced during this paper:

- Very few works were done on the presented topic.
- Work was vast and due to time limitation, it was not possible to conduct a full scale review
- Without experimental study, it was hard to explain the tools

Tools which were used for the study are of same segments and were hard to differentiate

## VIII. FUTURE SCOPE

This study can be used as a base for conducting experimental studies. Collecting large sample sizes will enhance and extend the limitations of the research which will help in attaining a proper outcome over the studied tools. More tools are there which can also be brought under study. To overcome the limitations in future researches, experimental study can also be conducted.

## REFERENCES

- [1] Prakash K. Ukhalkar, Dr. Rajesh N. Phursule, Dr Devendra P Gadekar, Dr Nilesh P Sable "Business Intelligence and Analytics: Challenges and Opportunities" International Journal of Advanced Science and Technology 2020.
- [2] Bhoomraogari Sai Teja "A Comprehensive Study on Business Analytics" International Journal of Engineering Research & Technology 2020.
- [3] Ifeyinwa Angela Ajah and Henry Friday Nweke "Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications" Big Data Cogn. Comput. 2019.
- [4] Valter Moreno, Felipe Elias Lobo Vieira da Silva, Rodrigo Ferreira, Fernando Filardi "Complementarity as a driver of value in business intelligence and analytics adoption processes" Iberoamerican Journal of Strategic Management 2019.
- [5] Ritu Ratra, Preeti Gulia "Big Data Tools and Techniques: A Roadmap for Predictive Analytics" IJEAT 2019.
- [6] Gloria J. Miller "Comparative Analysis of Big Data Analytics and BI Projects" IEEE 2018.
- [7] Dr. Venkatesh Naganathan "Comparative Analysis of Big Data, Big Data Analytics: Challenges and Trends" International Research Journal of Engineering and Technology 2018.
- [8] Ms. Komal "A Review Paper on Big Data Analytics Tools" IJTIMES 2018.

- [9] Althaf Rahaman.Sk, Sai Rajesh.K, .Girija Rani K “Challenging tools on Research Issues in Big Data Analytics” IJEDR 2018.
- [10] Emmanuel Ahishakiye, Elisha Opiyo Omulo, Danison Taremwa, Ruth Wario “Comparative Analysis of Open source Business Intelligence tools for Crime Data Analytics” IJLRET 2017.
- [11] Sridevi Bonthu, K Hima Bindu “Review of Leading Data Analytics Tools” International Journal of Engineering & Technology 2017.
- [12] Rachida F. Parks, Ravi Thambusamy “Understanding Business Analytics Success and Impact: A Qualitative Study” Information Systems Education Journal (ISEDJ) 2017.
- [13] D. P. Acharjya, Kauser Ahmed P “A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools” IJACSA 2016.
- [14] Moshen Marjani, Fariza Nasaruddin, Abdullah Gani, Ibrahim Abaker Targio Hashem, Aisha Siddiqa “Big IOT Data Analytics: Architecture Oppourtunities and Open Research Challages” IEEE 2017.
- [15] Nada Elgendy and Ahmed Elragal “Big Data Analytics: A Literature Review Paper” ICDM 2014.
- [16] Lekha R. Nair, Sujala D. Shetty “Research in Big Data and Analytics: An Overview” International Journal of Computer Applications 2014.
- [17] Prokopova Zdenka, Silhavy Petr, Silhavy Radek “Data analysis: tools and methods” Research Gate 2011.
- [18] Joaquim Lapa, Jorge Bernardino, Ana Figueiredo “A Comparative Analysis of Open Source Business Intelligence Platforms” ISDOC 2014.
- [19] Joseph O. Chan “An Architecture for Big Data Analytics” Communications of the IIMA 2013.
- [20] Mr. Mahesh G Hudda “A Survey on Big Data Analytical Tools” IJLTET 2013.