

Introduction to Tools for Data Analytics

15

SLOB Mapped against the Module:

To equip oneself with application-oriented knowledge of Tools for Data Analytics to facilitate management decisions for optimisation through resource allocation, managing competition, work scheduling and managing cost overrun, demand estimation, production and cost analysis etc.

Module Learning Objectives

After studying this module, the students will be able to:

- ⦿ Understand the meaning of Data Analytics.
- ⦿ Understand the steps of Data Analytics.
- ⦿ Know about the various types of Data Analytics.
- ⦿ Know about the different Tools used for the Data Analytics

Introduction to Tools for Data Analytics

15

Tools of Data Analytics actually refer to different software used for the purpose of Data Analytics. Before discussing about these, it is felt that a few words about Analytics is going to be helpful to understand the necessity of various types of tools related to it.

Analytics is the scientific process of discovering and communicating the meaningful patterns which can be found in data. In other words it is the science that analyses raw data to fetch information contained in it. It is concerned with turning raw data into insight for making better decisions. Analytics relies on the application of statistics, computer programming and operations research in order to quantify and gain insight to the meanings of data. It is especially useful in areas which record a lot of data or information. In today's information age, data is everywhere and ever increasing. As soon as internet is used by one (which may be from a desktop computer or laptop or tab or from a mobile phone etc.) digital footprint is left which gives rise to data of some kind.

Thus 'Data Analytics' is a broad term that encompasses many diverse types of Data Analysis. Any type of data or information can be subjected to the techniques of data analytics to get insight that can be used to improve things. These techniques can reveal trends and metrics that would otherwise be lost in the mass of information, The output of data analytics can then be used to optimize processes to increase the overall efficiency of a system, including that of a business. Implementing it into business models means companies can help reduce costs by identifying more efficient ways of doing business and by storing large amounts of data. Actually this helps to make better business decisions by analysing customer trends and level of satisfaction which can lead to development of new and better products and services.

So the role of Data Analytics for a business can be pointed out as follows –

- ⊙ **Gather hidden insights** – Hidden insights from data are gathered and then analysed with respect to the requirements of the business.
- ⊙ **Generate reports** – Reports are generated and passed on to the respective teams and individuals to deal with further actions for a high rise in business.
- ⊙ **Perform Market Analysis** – Market analysis is performed to understand the strengths and weaknesses of the competitors.
- ⊙ **Improve business requirements** – Analysis of data facilitates improvement in business by meeting the actual requirements of the customer.

Steps in Data Analytics

Step 1

The first step is to determine the data requirements or how the data is grouped. Data may be separated by age, demography, income, gender etc. Data values may be numerical or be divided by category.

Step 2

The second step in Data Analytics is the process of collecting it. This can be done through a variety of sources – computers, on line sources, cameras, satellites or through personnel.

Step 3

Once the data is collected, it must be organized so that it can be analysed. This may take place on a Spread-sheet or other form of software that can take statistical data.

Step 4

The data is then cleaned up before analysis. This means it is scrubbed and checked to ensure there is no duplication or error and that it is not incomplete. This step helps correct any error before being analysed. This process is also known as Data Wrangling.

Step 5

Analyse data and generate report.

Types of Data Analytics

There are four types of Data Analytics depending on the type of data available and the type of knowledge one is looking for from the available data. These are –

- 1. Descriptive Analytics:** This is the most basic type of Analytics. It looks at data to examine, understand and describe something that has already happened. It provides quantitative information on “What happened” by analysing the data with the help of statistical techniques like calculation of Mean, Median, Mode etc. Daily report of Cost of Production of an item being produced continuously in an automatic machine is an example.
- 2. Diagnostic Analytics:** It goes deeper than Descriptive Analytics by seeking to understand the “Why” behind what happened. While the production report may show an increase in the Cost of Production of the item, but the reason behind the same is not mentioned. Diagnostic Analytics will look into the patterns and deviations to find the root cause of increase in the cost. Diagnostic Analytics involve correlation of two different datasets.
- 3. Predictive Analytics:** This relies on historical data, past trends and assumptions to answer questions about “What will happen” in future. This is of immense importance as far as the situations of future prediction (like Sales forecasting) are concerned. Predictive Analytics correlates the results of Descriptive and Diagnostic Analytics with external datasets.
- 4. Prescriptive Analytics:** This identifies the specific actions an individual or organisation should take to reach future targets or goals. In other words it comes up with recommendations of “What actions should be taken”. In fact different possible scenarios are taken into account and the corresponding outcomes of each is considered of which the best possible one is recommended. Huge computing power is necessary for going through such type of analytics. This type of Analytics skill is observed in Artificial Intelligence (AI) and Machine Learning (ML).

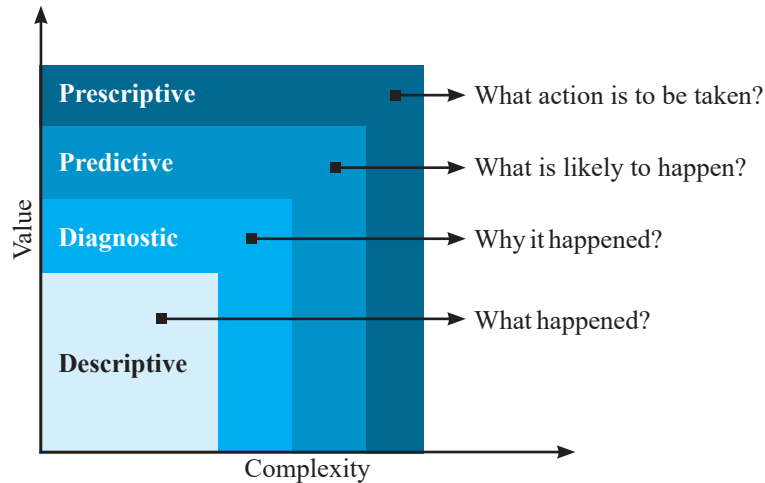


Figure 15.1: Types of Data Analytics

Besides the above four, there can be other type known as –

Exploratory and Confirmatory Data Analytics

As the name implies, Exploratory Data Analytics explores the dataset to find answer as to which describes the data well. In this case no hypothesis is stated before analysing the data. It starts by framing questions about what to do with the data. Thereafter the data is manipulated to the best possible extent to find the answers to the questions. This involves understanding data structure, identifying erroneous or missing data, establishing a margin of error and figuring out the hypothesis.

Confirmatory Data Analytics deals with some stated hypothesis. Same is tested for the dataset using statistical tools and inference is drawn about its correctness. An example of hypothesis can be – “I predict that a person’s likelihood of recommending our product is directly proportional to their reported satisfaction with the product”. Now if the trend analysis is carried out on the available data then from that the hypothesis is either proved or disproved. But the availability of the stated hypothesis gives a clear cut direction towards which the data analysis is done. In a way it is obviously advantageous compared to the Exploratory Data Analytics.

Tools for Data Analytics

From the above discussions, it is clear that acquiring and storing data, processing the same and finally report generation are the steps in Data Analytics in a nutshell. For these, the tools used are nothing but different Software. Depending on the purpose for which a tool is used, one can classify them as follows.

1. Data Transformation Tools

These are basically used for the purpose of cleansing of the available data. A prime requirement of getting meaningful output from any software is to provide it with the data in specific field types and structures. Normally lot of issues are noticed in the datasets as far as formats, structures, spellings etc. are concerned. To take care of these, **ETL (Extract, Transform and Load)** tools are used. It can be mentioned here that according to some other line of thinking, the tools are **ELT (Extract, Load and Transform)** type and not ETL type which works better. As far as our purpose is concerned both work perfectly alright.

Such tools facilitate building a pipeline through which the data are brought together from multiple sources,

cleansed and stored in a place called data warehouse for analysing. This is applicable when the data is structured. In case of unstructured or semi structured data, the storage is called Lake instead of Warehouse.

Data Stores filling start by data extraction from primary sources. Various data inconsistency is removed during this phase. Before their transformation into the data schema, extracted data can be loaded in a temporary dumping ground known as Data Staging Area (DSA). The data component in DSA used to be most frequently a part of those solutions of data stores which has a source in heavy transaction systems. Use of DSA will reduce requirement of transaction system's utilization in the ETL Process. Use of DSA is possible also in the case when it is necessary to transfer data from a text file into the required database format. After the extraction follows data transformation which will convert data obtained from single data source into unified data model. This model makes it possible to create aggregation and clustering.

The final phase of ETL is data transmission from source data memories or temporary dumping ground to database tables of the data store. At the primary filling it can be a gigantic quantity of data. Because ETL works in batch mode, such amount of data will be brought again at the time of next regular updating. The time interval of updating could be a day, a week, a month or a year.

The transformed data coming from various service systems and external databases are saved in Data Warehouse which is an extensive central business database. Actually the data in the Data Warehouses are not optimized for quick transaction processing but meant for quick administration of analytical information obtained from big amount of data. When the concept of decentralization is applied to Data Warehouse then we get Data Marts. Basically these are decentralized, based on certain themes. Data Marts are meant to provide analytical information to specific groups like Marketing, Sales, Finance etc.

Data Transformation Tools are also called Data Smoothing Tools.

Many software for visualization and analysis have such ETL (or ELT) tools built in. Besides, there are many special tools which are capable of handling this cleansing process.

Xplenty, Stitch, ABS Glue, Skyvia etc. are some of the commonly used ETL Tool for data cleansing.

2. Data Analysis Tools

These tools are meant for Analysis part of the data. Broad classification of these tools are –

- Spread-sheets
- Business Intelligence (BI) tools
- Financial data Analytics tools
- Programming Languages
- Tools for Statistical Data Analysis
- Industry Specific tools

I. Spread-sheets

Although there are several Spread-sheet database software solutions that offer more than a traditional spread-sheet, Spread-sheet software is still indispensable for many businesses. From generating reports and creating a budget to becoming a glorified to-do list, spread-sheets have the flexibility to meet just about every need of user. The right Spread-sheet software can not only help the user to shift through thousands of datasets, but also keeps the user on track for setting and completing goals.

Of the several available Spread-sheet tools of today, **MS-Excel** remains the most common. The simplest and most obtainable analysis proceedings of business data is offered by MS-Excel. Certainly it is the cheapest

option available because there is no user having MS-Excel not installed in the computer system. So there is no necessity to buy license for specialized software. Users can straightway go for creation of analytical reports and graphs. Data analyses created by MS-Excel are very dynamic and effective. They enable a lot of different views and graphical representations. Data feeding to MS-Excel is possible in different ways. Most common is the manual data feeding from business reports. The other way is data import from business information system. The third way is direct connection to database of business information system. This way is most operative in the situations where huge volume of data is handled.

Pivot Tables are one of the most powerful tools of MS-Excel This enables data summarization, filtration and ordering. It is possible to create a lot of different views, reports and graphs from one data source. Created Pivot Table is easily editable – we can add or delete data, columns, rows or change summaries without the influence of Data Source.

Some of the reasons why MS-Excel finds its utility as a tool used in data science are as follows -

1. **Awareness:** Irrespective of the demographic background, MS-Excel is a widely known tool to almost all people who work with data, and prepare reports. The tables, pivots, VBA-scripts and charts are having high awareness. MS-Excel is treated as a personal productivity tool used extensively in business organisations for preparing dashboards and reporting.
2. **User-friendliness:** Microsoft Excel is incredibly easy to use. Especially for beginner data scientists, the ease of use and the wide-scale availability of Microsoft Excel makes it an excellent introductory program for both students and professionals.
3. **Data visualization:** MS Excel is one of best editors for data visualization in 2D format. The tables are easily edited, formatted, coloured, and shared. Google Sheets is a clear confirmation of the Excel design for editing data — but scaled for multiple users. Add to that the built-in tables, filters, slicers, groupings, window splitting, cell formulas and other features and any alternative table or database editor set MS Excel apart from many tools. MS Excel can handle as many as 1,048,576 rows and 16,384 columns per sheet. Therefore, other than a set of complex real-life problems dealing with truly big data, MS-Excel can solve the purpose in many cases. MS Excel also helps in creating the metadata and also to change metadata over time as a dataset grows.
4. **Data Analysis:** Microsoft Excel has multiple functions programmed into the software which makes it simple to explore a dataset through sorting, filtering, and pivot tables. Tools like Scenario Manager also make it easy to create and compare hypothetical scenarios across datasets. Moreover, MS Excel helps in unfolding the pattern and psychology of the data. For a beginner in data science, MS Excel is a very useful tool. Further, Excel comes with various analytical tool packs that can be activated to unleash more advanced calculations. In addition, one can write customized scripts vis-à-vis the algorithm in use, for data analysis. Even there is an option for using Python with MS Excel.

Comparability with other software: This is one of the very powerful features of MS Excel as a tool. The other tools like SPSS, JAMOVI, R, Python etc directly allow to import the excel sheet (in CSV form) as the input data for further analysis. Data science professionals that are committed to writing code and queries can use SQL.

Besides MS-Excel there are other Spread-sheet software available for Data Analytics and some of those are – **Google Sheets, Quip** etc.

Google Sheets is a free alternative to Excel. It requires a Google account to begin work (which one will

readily have if the person is user of gmail). It offers real-time collaboration, commenting and version history to accurately track all progress between contributors. Google Sheets can immediately sync with G Suite apps like Google Analytics and Google Data Studio. It can also populate a spread-sheet with data pulled from Google Forms. In addition to its collaborative features, Google Sheets also comes equipped with a slew of traditional mathematical and logical functions. Even better, its available in the web, Android and IOS Systems, making it one of the most easily accessible software options. Google Sheets and all of its features are available up to 15 GB storage.

Quip is a unique productivity tool that combines spread-sheets, documents and team chat all into one app. The tool features the flexibility and customization. For instance you can toggle between document and spread-sheet layout as you work. Quip also boasts more than 400 built-in functions and offers key board shortcuts for most menu items, such as (Shift + Space) for row selection and (Ctrl + Space) for column selection. Plus just like Excel, creating a Graph or Chart is as easy as selecting a data range and clicking a button (in this case, Insert and Chart)

II. Business Intelligence (BI) Tools

• Meaning of BI Software

Business Intelligence (BI) Software is a set of business analytics solutions used by companies to retrieve, analyse and transform data into useful business insights usually within easy-to-read visualization - like charts, graphs and dashboards. Examples of the best BI Tools include data visualization, data warehouses, interactive dashboards and BI reporting tools. A BI Solution pulls internal data produced by a company, into an Analytics platform for deep insights as to how different parts of a business affect one another.

As Big Data has gained in prominence, the tendency for companies to collect, store and mine their business data has increased many times and so has the popularity of BI Software. Companies generate, track and compile business data at a scale never seen before. The ability to integrate cloud software directly with proprietary systems has further driven the need to combine multiple data sources and take advantage of data preparation tools. But all this data is nothing if we can't make sense of it and use it to improve business outcomes.









To make informed choices, businesses need to make their decisions on evidence. The mountains of data that businesses and their customers produce contain evidence of purchasing patterns and market trends. By aggregating, standardising and analysing that data, businesses can better understand their customers, better forecast their revenue growth and better protect themselves against business pitfalls.

Business intelligence has traditionally taken the form of quarterly or yearly reports that report on a defined set of Key Performance Indicators (KPI). But today's BI Reporting software is backed by Data Analytics tools that work continuously at the speed of light. These insights can help a company take a course of action within minutes.

BI Software interprets a sea of quantifiable customer and business actions and returns queries based on patterns in the data. BI comes in many forms and spans, many different types of technology.

The chart below shows a comparison of few top Business Intelligence Tools according to user popularity and major features.

Table 16.1: Chart showing comparison of top BI Tools [Source – www.technologyadvice.com]

Product	TA Rating	Data Analytics	Natural Language Processing	Real-Time Reporting	Embedded Analytics
	4/5	✓	⊘	⊘	✓
	4.35/5	✓	⊘	✓	✓
	4.5/5	✓	✓	✓	✓
	4.5/5	⊘	⊘	✓	⊘
	4.5/5	✓	✓	✓	✓
	4/5	✓	⊘	✓	⊘
	4.5/5	✓	⊘	⊘	✓
	4/5	✓	⊘	✓	⊘

In the chart TA Rating means the rating given by Technology Advice

- **Data storage for BI**

Throughout an organisation data is available in a number of systems. Companies should ensure standardised formatting across data types from each of these systems so that most accurate analysis is possible. A large enterprise can have information about the customers in the Customer Relationship Management (CRM) application and have financial data in the Enterprise Resource Planning (ERP) application and several other key revenue datasets in various cloud software applications. These separate programs may label and categorise data differently and the company will need to standardise the data before analysis.

Some BI Platforms pull data for analysis directly from the source applications through a native API connection or Webhook. Other BI Tools require the use of a cloud data storage system to aggregate diverse data sets in a common location. Small businesses, single departments or individual users may find that a native connection works well, but big organisations which generate large datasets will need a more comprehensive BI setup.

In case they want to go for centralised storage location then a Data Warehouse or Data Mart could be a

solution and purchase an ETL Tool to facilitate their Big Data storage. Alternatively they may use a Data Storage Framework like Hadoop to manage their data.

- **Analysing Big Data with BI**

Regardless of whether businesses choose to store their data in a Data Warehouse, a Cloud Database, an On-premise Server or run queries on the Source System, data analysis and the resulting insights make the field appealing to the business users. Data Analytics tools vary in terms of complexity, but the general method of combining large amounts of normalised data to identify patterns remains consistent across BI Platforms.

Data Mining:

This is the activity of “Data Discovery” because here the patterns and inconsistencies of data unveiled through automated or semi-automated data analysis. Common correlations drawn from Data Mining include grouping specific sets of data, finding outliers in data and drawing connections and dependencies from disparate datasets.

Data Mining often uncovers the patterns used in more complex analyses, like Predictive modelling which makes it an essential part of the BI Process whose growth is correlated directly with the rise of Big Data in businesses of all sizes.

Of the standard processes performed by Data Mining, association rule learning presents the greatest benefit. By examining data to draw dependencies and construct correlations, the association rule can help businesses better understand the way customers interact with their website or even what factors influence their purchasing behaviour.

Association rule learning was originally introduced to uncover connections between purchase data recorded in point of sale systems at supermarkets. For example if a customer purchased Tomato Sauce and Cheese, the association rules would likely uncover that the customer purchased Hamburger Meat as well. Though this is a very simple example but it works well to understand the type of analysis that now connects incredibly complex chains of events in all sorts of industries and helps users find correlations that would have remained hidden otherwise.

Predictive Analytics with BI Software:

One of the most exciting aspects of BI, advanced analytics features like Predictive and Prescriptive analytics function as a subset of Data Mining. The tools use existing datasets and algorithmic models to help companies make better decisions.

As discussed before, Predictive Analytics forecast future events based on current and past data. By drawing connections between datasets, these software applications predict the likelihood of future events which can lead to a huge competitive advantage for businesses.

Predictive analysis involves detailed modelling. Even it ventures into the realms of Artificial Intelligence (AI) and Machine Learning (ML) where software actually learns from past events to predict future consequences. The three main forms of Predictive analysis are Predictive Modelling, Descriptive Modelling and Decision Analytics.

Predictive Modelling is the most well-known segment of Predictive Analytics. This type of software predicts, particularly in reference to a single element. Predictive models use algorithms to search for correlation between a particular unit of measurement and at least one or more features pertaining to that unit. The goal is to find same correlation across different datasets. An example of this modelling is “Likelihood of a customer switching Insurance Providers”.

Descriptive Modelling seeks to reduce data into manageable sizes and groupings. Descriptive Analytics works well to summarise information such as unique page views or social media mentions.

Decision or Prescriptive Analytics take into account all the factors related to a discrete decision. It predicts the cascading effect an action will have across all the variables involved in making that decision. In other words Decision Analytics gives businesses the concrete information they need to predict outcomes and take action.

Natural Language Processing:

Data comes in three forms – structured, semi structured and unstructured. Of these, the most common is unstructured data which includes text documents and other types of files that exists in formats which computer cannot read easily.

Unstructured data cannot be stored in neatly categorised sets of similarly formatted data rows or columns. This makes the traditional data mining software unable to analyse the data. But often such data is seen to become crucial for understanding business outcomes. With so much data in unstructured form, Text Analytics should be a key consideration while deciding on the purchase of BI Tool.

Natural Language Processing (NLP) software, also known as Text Analytics software combs large sets of unstructured data to find hidden patterns. NLP is particularly interesting for businesses that work with social media. Using the right software mix of data ingestion and AI, a business can set up rules to track key words or phrases – for example “Name of a business” – to find patterns in how customers use that language. NLP Tools also measure customer sentiment, provide actionable insight into lifetime customer value and learn customer trends that can inform future product lines.

- **BI Software for Corporate reporting**

The previous two applications of BI Software dealt with the mechanics of BI System – how business data is stored and how software refines this data into meaningful intelligence. BI reporting focuses on the presentation of these findings.

Online Analytical Processing (OLAP):

This uses multidimensional databases to enable users to query data warehouses and create reports that view data from multiple perspectives. OLAP gives BI Software the ability to combine data, drill down into single metrics and view data for combinations of single metrics that are unobtainable in a traditional spread-sheet setup.

For example, a supply chain’s data metrics can include Location, SKU, Date of purchase, Name of the Salesperson and Expiration date. OLAP Tools can provide the analysts with the power to surface insights that would otherwise be hidden within two or three dimensional spread-sheets.

Data Visualisation:

One of the most popular trends of BI is Data Visualisation. This allows companies to graphically the results of data mining or other analytics. Presenting findings in a visual format like – a graph, chart or on a map provides immediate insights into the most important metrics – awareness that does not surface within the context of a spread-sheet.

Dashboards:

Every user of a system does not need full access to everything available in the dashboard. Most employees

only need access to a dashboard of their most important metrics. It gives at a glance access to a range of predefined visualisations. While each company can define its own dashboards based on custom business needs, some of the possible dashboards that BI Tools provide are as follows –

- Sales Dashboards that includes the total number of leads and prospects in each stage of the sales funnel, KPI metrics of the total number of meetings scheduled per salesperson, a total revenue leader-board, gas gauge tool that shows total revenue towards monthly goal
- Marketing Dashboards that shows a line chart with the total number of marketing qualified leads per day, top performing blog posts per month, latest social post
- Customer Success Dashboards with visualisations for the total number of open tickets, number of closed tickets per day, average time to close, ticket totals leader-board
- IT Dashboards with key metrics regarding sprint progress, total number of open bug tickets, current on-call developers, feature request leader-board

Some BI solutions also offer interactive dashboard tools where business users can manipulate the data visualisations, dig for a more detailed view and zoom out for more context.

Alerts and Notifications:

While interactive dashboards and reports greatly extend the usability of BI Software for non IT users, alerts and notifications can provide even further practical applications for all business users. Alerts notify users who do not spend most of their time in the tool to data changes that need immediate action.

When the companies set alerts for threshold of high and low performance, they can track when they need to mobilize a response or investigate an issue before it becomes an emergency. Even better, companies that set alerts for goal metrics can celebrate and recognise their team efforts early and often.

Embedded Analytics:

BI applications promise to clarify data analytics tools for the most non-technical of employees, which has driven the demand for Embedded BI. These features let companies build data visualisations within their cloud BI Software and dynamically serve those visualisations to internal and external customers within company apps.

Visualisations, reports and dashboards that are embedded in a company webpage or cloud app save companies thousands of hours and lots of money they would otherwise use to build BI Reporting tools and Analytics dashboards from scratch to track company's performance. These tools now give the business users access to custom, plug-and-plug visualisations, greatly speeding the time to market.

- **Different BI Software from various vendors**

Business Intelligence Tools and Platforms come in several forms for varying business needs. Companies looking to provide data services to business users will find Self Service BI meeting the needs of most of their users. Data Visualisation Tools are helpful for teams that are dipping their toes into Data Analytics but may not have lots of extra development resources available. Data Warehousing Tools provide the underlying infrastructure that can house and cleanse data before serving it up through visualisations and BI Tools provide end-to-end dashboard tools to store, cleanse, visualise and publish data. A list of top BI Tool Software Vendors is given below:

Table 16.2

Self Service	Data Visualisation	Data Warehousing	Bi Platforms
SAP Crystal Reports	iDashboards	Sisense	Tableau
Chartio	Dundas	Oracle BI	InsightSquared
Alteryx	Segment	SAS	Domo
Jaspersoft	Geckoboard	Birst	SAP Lumira (formerly Business Objects)

Some of the important Business Intelligence tools are discussed in the following paragraphs.

Tableau specialises in making beautiful visualisations. It is a very advanced tool and one of the early entrants in the market. It is considered to be the industry leaders among BI Tools. Tableau's platform certainly live up to the hype. In addition to easy-to-use Reporting, a full API, Report Sharing and good customer support are the plus points in support of its purchase. But it is quite costly and affordable for corporate business houses. That's why much of their advertising is focussed on corporate environments with data engineers and bigger budgets. Tableau has a public (free) version but with limited capabilities. The more you pay the more you can access with Tableau including benchmarked data from third parties.

Tableau allows to pull data from an array of cloud software and data preparation sources, including but not limited to Excel, SQL and Oracle.

InsightSquared is built specifically to help with Sales, Marketing, Finance, Staffing and Support Services. It is perfect for detailed insights about specific aspects of a company's operations. In addition to data analysis and reports InsightSquared also allows one to get an idea about Closed business deals, Sales success rates of certain salesperson and more.

Integrating InsightSquared with Salesforce gives users access to predictions about which prospects are most likely to lead to a sale, which salespeople are likely to win most sales and more.

Domo is the first BI Tool to run completely on the cloud. Domo is a quickly growing system that allows the user to access insights about the business from anywhere. Domo offers over 500 data connectors and its own app store allows the user to plug in apps specifically tailored for the industry to which the user's business belong.

Names of few more important BI Softwares are – **QlikView, Birst, GoodData, Hubble, Looker, MicroStrategy, BOARD, Microsoft PowerBI, Oracle BI, Logi Analytics, Sisense, Alteryx** etc.

III. Tools for Financial Data Analytics

Harnessing the power of technology, Financial Analytics tools are transforming business operations and planning. Financial Analytics, a subset of Business Intelligence (BI) and Enterprise Performance Management (EPM), has become indispensable in the modern business environment. Using financial analytics tools companies can examine huge volumes of financial and other data to identify patterns and predict trends, boost revenues and slash costs.

Financial Analytics involves the use of software applications to examine all the data generated by an organisation. This includes Financial and Accounting information as well as other data like capturing customer interactions, monitoring supply chains, deep diving into historical trends and more.

As the complexity of business world has increased, companies need to deploy more sophisticated tools to understand what is going on and to plan. Financial Analytics improves on traditional financial analysis

because of its ability to capture larger volumes of data as well as to drill down to a highly detailed level. Financial Analytics does not only reveal historical trends, but using Predictive Analytics it can peer into the future. Answers to the following critical issues can be provided by the Financial Analytics tools.

- ⦿ What is the profitability of different products or services?
- ⦿ What are the most valuable customer segments?
- ⦿ What risks are the business currently facing and what are the potential future challenges?
- ⦿ How are different sales channels performing?
- ⦿ What future events could impact stock price performance?

Types of Financial Analytics

Financial Analytics can provide business leaders a better understanding of business processes and drivers. This grants them the ability to monitor and improve business operations and to plan effectively for future performance. Different types of Financial Analytics are –

- Predictive Sales Analysis
- Product Profitability Analytics
- Value Driver Analytics
- Financial Ratio Analytics
- Scenario and Sensitivity Analysis
- Growth Rate Analysis
- Cash Flow Analysis
- Variance Analysis

- **Predictive Sales Analysis:**

This enables an organisation to assess and adjust Sales Forecasts more rapidly than when using Excel based models. Predictive Sales Analysis can incorporate many more data points to predict future sales because of the ability to drill down deeper, for example, into the performance of different sales channels and different product lines.

- **Product Profitability Analytics:**

A company's overall profit is the money left at the end of an accounting period after subtracting total costs from the total revenue. Product Profitability Analysis is used to determine the profitability of each product or service. This is an invaluable insight for decision makers as the profitability of different product lines can be masked by factors such as high sales volume or high revenue growth. Product revenues are stripped off all associated costs – production costs, logistics, discounts, rebates, marketing and commissions – to reveal the profit derived from each product.

Business leaders need to know how to allocate their resources in the most effective way – finding out which products are most profitable is critical.

- **Value Driver Analytics:**

In business Value Driver are the elements that increase the worth of a product, service, asset or business. In the case of a product it could be a differentiating capability that makes the product a “Must have” for customers. For a business it could be a trusted staff or a customer base which increase the value of the business.

Most businesses have identified their main value drivers as part of their strategy development. What value driver analytics provides is a way to quantify, measure and assess these drivers using transparent financial metrics.

- **Financial Ratio Analytics:**

Ratio Analysis is simply the analysis of different financial metrics in a business' financial statements. Financial Ratio Analysis compares the relationship or ratio between two or more financial items. Financial ratios are used in six main areas – Liquidity, Coverage, Solvency, Profitability, Efficiency and Market Prospects.

Analytics allows us not only to see different ratios more quickly and accurately than in the past but also let Data Analysts visualise this data in more understandable formats, providing managers with the evidence they need to make informed decisions.

- **Scenario and Sensitivity Analysis:**

Business leaders make business decisions based on an analysis of risks and benefits. Any decision is based on a set of assumptions – i.e market and economy. Scenario and Sensitivity analysis model these assumptions and manipulate them to show the impact of different decisions or different conditions. Scenario Analysis examines the results of all the assumptions – while Sensitivity Analysis examines the impact of changing a single variable. This can include macro or micro economic changes. The impact of pricing and volume changes can also be examined.

Analytics takes this to another level by allowing for the development of more complex models that can be updated more rapidly to reveal which factors the business is most sensitive to – such as inflation or customer confidence.

- **Growth Rate Analysis:**

Growth Rates are the most common financial metrics used in business. Simply put, growth rates are the percentage change of a specific variable within a given time period. They are used by business executives to track performance. Businesses typically track revenue and profitability growth, sales and cost evolution and so on. Investors and financial analysts typically use growth rate analysis to assess potential investments including revenues and earnings, price to earnings ratio (P/E), price to earnings to growth ratio (PEG) and return on equity (ROE).

- **Cash flow Analysis:**

Cash is most important resource in business. Being able to determine at any point in time how much cash a company has or will have in a specific future period is essential. Tracking cash flow is perhaps the most important function performed by the finance department. Cash flow analysis reveals how much money is coming into the business from sales, services or investments or how much money is going out in the form of costs and interest payments. Being able to control and predict these flows determines the viability of any business. Cash flow is divided into –

- (1) Cash flow from operating activities,
- (2) Cash flow from investing activities and
- (3) Cash flow from financing activities.

- **Variance Analysis:**

Variance Analysis is an important aspect of budgeting activities. At the most basic level, we use it to

compare actual performance to what was predicted in company forecasts or budgets. In other words it compares actual results with planned one.

Variance Analysis can highlight various issues. Calculating and analysing variances can identify areas where management needs to put more focus – for example in optimising costs or increasing operational efficiency. It can also point out the flaws in the budgeting or planning process if there is significant or persistent variance between the actual and planned.

Data analytics is a thrilling development in financial analysis because many of these analytics can now be performed much more quickly and accurately. Business Intelligence and Automation Platform tools now help us to extract, organise and examine data from different sources and systems. This provides invaluable support to the company executives – allowing them to make faster, more informed and better decisions.

Financial Analytics Tools

These are mainly BI Tools with option for Financial Analytics. Some of the BI applications with specific modules for Financial Analytics are given below.

Jedox – Salient features of this tool are as follows.

- ⦿ SAAS based solution Business Intelligence within seamlessly integrated software platform.
- ⦿ Deployable in the cloud as well as on-premise server or hybrid
- ⦿ Facilitates business planning, reporting, data consolidation and financial data analytics
- ⦿ Offers Excel add-in enabling data modelling, ad-hoc and predictive analytics, KPI reporting, budgeting, forecasting activities within the familiar Excel environment.

Zoho Analytics – Following are the salient features of this tool

- ⦿ BI and Analytics platform
- ⦿ Deployable in the cloud as well as on-premises server
- ⦿ Can be used across various functional areas and by a wide range of users for their reporting and analytics needs
- ⦿ Allows users to easily create and share powerful reports within minutes
- ⦿ Current version of Zoho Analytics has additional features like Zoho DataPrep., a Unified Data Management and Analytics Platform, Deep Augmented capabilities like Zia insights and Conversational BI, Data Story telling features

Quickbooks – This has the salient features given as –

- ⦿ One of the most feature rich accounting solutions for small businesses
- ⦿ User friendly and accessible platform, providing all the tools needed for day to day financial management.
- ⦿ Ability to track and record business information at high level of detail – providing treasure trove of “Small” Data.
- ⦿ Can generate reports but requires integration with other data analysis/ reporting tools to provide more advanced Financial Analytics.

Few other tools for Financial Data Analytics are **Hyper Anna, NetSuite, FICO** etc.

IV. Programming Languages

Data scientists use several Programming Languages in their work. Programming allows the creation of

specific analytical solutions which may not be available in other ready to use software packages. Some of the most popular languages for this purpose are – R Programming, Python, SAS etc.

- **R Programming :**

R is a open source programming language that is widely used as a statistical software and data analysis tool. It generally comes with command line interface. R is available across widely used platforms like Windows, Linux, Mac-OS. It is the latest cutting edge tool. Due to its expressive syntax and easy-to-use interface, it has grown in popularity in recent years. In fact this is one of the most popular languages amongst statisticians, data analysts, researchers and marketers to retrieve, clean, analyse, visualize and present data.

It was designed by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand and currently developed by the R Development Core team. R programming language is an implementation of the S programming language. The project was conceived in 1992, with an initial version released in 1995 and a stable beta version in 2000.

Reasons for using R

The figure below is meant for showing the reasons of using R. The points mentioned are discussed briefly in the paragraphs next.



Figure 15.2: Reasons for using R

- ⊙ It's an Open Source language. It can be installed for free because it is licensed under the GNU General Public License. There are many R Packages available under the same license. So there is no need to pay any license fee for the usage of any of these packages – even for commercial applications.
- ⊙ R Programming is Platform Independent language. Actually it is compatible with all the popularly used Operating Systems like Windows, Linux and Mac-OS. R code written on one platform can easily be ported to another without any issues. Cross platform interoperability is an important feature to have in today's computing world – even Microsoft is making its coveted .NET Platform available on all platforms after realising the benefits of technology that runs on all platforms.

- ⦿ It is used as a leading tool for Machine Learning, Statistics and Data Analysis. R can easily create objects, functions and packages based on user's requirements.
- ⦿ It allows the user to integrate it with other languages like C, C++ etc. As a result, easy interaction with many data sources and statistical packages is possible.
- ⦿ R is currently the most wanted programming language in the Data Science jobs. Hence it is considered to have the hottest trend in the job market.
- ⦿ The R Programming language has a vast community of users and it's growing day by day. It is being used by the biggest tech giants which is a sign of the potential of the language. R has the right mix of power and simplicity. Here are a few ways industry stalwarts are using R and contributing to the R ecosystem.

COMPANY	APPLICATION / CONTRIBUTION
Twitter	Monitor user experience
Ford	Analyse social media to support design decisions for the cars.
New York Times	Info-graphics, Data journalism
Microsoft	Released Microsoft R Open, an enhanced R distribution and Microsoft R Server after acquiring Revolution Analytics.
Human Rights Data Analysis Group	Measure the impact of war
Google	Created the R style guide for the R user community within Google

- ⦿ R is currently the most wanted programming language in the Data Science jobs. Hence it is considered to have the hottest trend in the job market.

- **Applications of R Programming in the real world**

1. **Data Science** – With the advent of “Internet of things” (IoT) devices creating terabytes and terabytes of data that can be used to make better decisions, Data Science is a field that has no other way but to go up. Simply explained, a data scientist is a statistician with an extra asset – computer programming skills. Programming languages like R give a data scientist superpowers that allow them to collect data in real time, perform statistical and predictive analysis, create visualisation and communicate actionable results to the stakeholders.
2. **Statistical Computing** – R is the most popular programming language among statisticians. In fact it was initially built by statisticians for carry work related to statistical data. It has a rich package repository with over 9000 packages having every statistical function one can think of. R's expressive syntax allows researchers – even those from non-computer science backgrounds to quickly import, clean and analyse data from various data sources. R also has charting capabilities which means one can plot the data and create interesting visualisations from any dataset.
3. **Machine Learning** – R has found a lot of use in predictive analytics and machine learning. It has various packages for common ML tasks like linear and non linear regression, decision trees, linear and non-linear classification and many more. Everyone from machine learning enthusiasts to researchers use R to implement machine learning algorithms in fields like finance, genetics research, retail, marketing and health care.

- **Features of R Programming language**

1. **Statistical features of R:-**

- ⊙ Basic Statistics – Most common terms of basic statistics are Mean, Median and Mode which are the Measures of Central Tendency for a dataset. These can be very easily computed using R.
- ⊙ Static Graphics – R is rich with facilities for creating and developing interesting Static Graphics. R contains functionality for many plot types including graphic maps, mosaic plots, bi-plots and the list goes on.
- ⊙ Probability Distributions – Probability Distributions play vital role in statistics. By using R various types of problems related to probability distributions (such as Binomial Distribution, Normal Distribution, Student's t Distribution, Chi Square Distribution etc.) can be handled very easily.
- ⊙ Data Analysis – It provides a large, coherent and integrated collection of tools for data analysis.

2. **Programming features of R:-**

- ⊙ R Packages – One of the major features of R is the fact that it has a wide availability of libraries. R has CRAN (Comprehensive R Archive Network) which is repository holding more than 10,000 packages.
- ⊙ Distributed Computing – Distributed computing is a model in which components of a software system are shared among multiple computers to improve efficiency and performance. Packages like ddR and multidplyr are used for distributed programming in R.

- **Programming in R**

Since R is much similar to other widely used languages syntactically, it is easier to code and learn in R. Programs can be written in R in any of the widely used IDE like **R Studio, Rattle, Tinn-R** etc. After writing the program, the file needs to be saved with the extension **.r**

To run the program following command has to be used in the command line.

```
R file_name.r
```

Some examples of programs in R language are given as follows.

Example

Write R program to generate 10,000 numbers in a Random distribution, organise them based on the frequency and create a Bar Chart.

Solution:

Following three lines of code need to be written for the purpose

```
n <- floor(rnorm(10000, 500, 100))
t <- table(n)
barplot(t)
```

Thereafter these are to be copied into the RStudio window. In fact all the lines can be selected together by pressing (Ctrl+A) and then pressing (Ctrl+Enter) same can be copied into the RStudio screen. This is applicable if the operating system is Windows. In case of Mac-OS the commands are (Cmd+A) and (Cmd+Enter) respectively.

The output generated will be as follows –

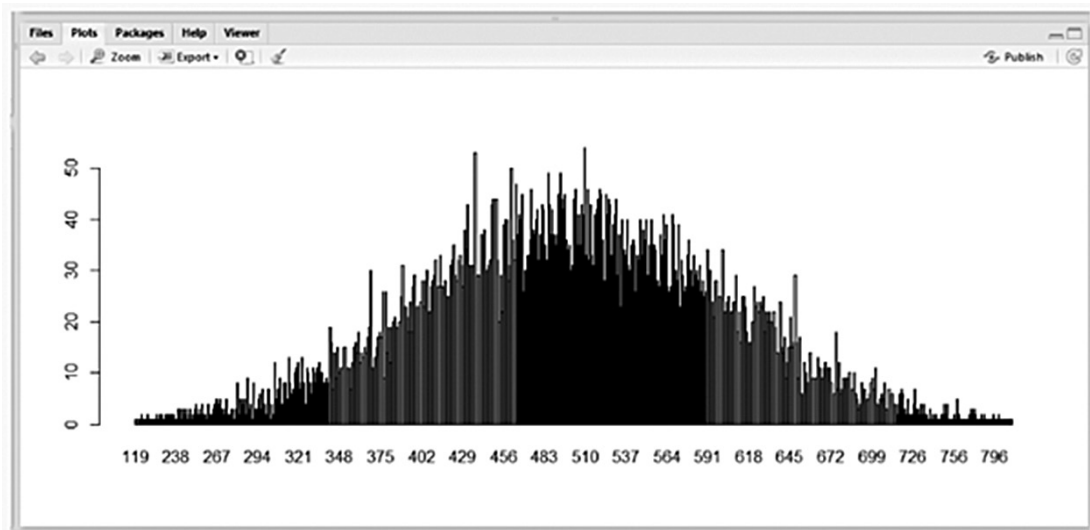


Figure 15.3: R studio screen showing the Bar Chart

This can be seen at the right bottom section of RStudio screen. From the plot it is clear that the data follows a Normal Distribution because it has taken a bell shape.

Explanation

The first line of code, that is: `n <- floor(rnorm(10000, 500, 100))` actually generates 10000 random numbers such that the Mean is 500 and Standard Deviation is 100.

The “floor” function takes each number in this list and removes the decimal points.

If this line of code is tried separately in the R console then the following type of output can be obtained.

```
> n <- floor(rnorm(10000, 500, 100))
> n
 [1] 392 554 575 357 516 521 499 734 424 457 724 549 338 600 564 415 349 454 416 615 384 342
[23] 533 545 491 396 625 492 377 603 583 536 486 616 675 627 602 459 476 390 455 450 428 785
[45] 481 514 482 386 515 423 519 638 539 457 432 363 558 564 491 511 509 513 409 421 571 682
[67] 419 739 295 254 547 532 513 394 438 652 508 425 344 488 598 496 350 438 394 415 428 644
[89] 313 669 723 505 553 231 520 576 383 516 306 539 597 437 482 298 535 475 644 493 389 310
[111] 201 402 449 526 664 431 452 551 479 469 513 734 448 459 554 581 616 513 519 487 413 520
[133] 481 692 556 497 510 347 368 526 396 534 508 425 406 483 494 435 482 634 491 535 600 487
```

Figure 15.4

[Note – Instead of showing all the 10000 numbers only a part of the R console screen has been shown here]

Next the “table” function considers all of these 10000 numbers and counts frequency of each number. The output of the same is as follows.

```

> t <- table(n)
> t
n
 60 132 140 150 160 164 165 170 174 176 182 183 185 187 188 192 193 194 195 196 197 198 199 200
 1  1  1  1  1  1  1  1  2  1  1  1  1  1  2  1  1  1  1  1  1  1  1  1  1
201 205 209 211 214 215 216 217 219 220 221 222 226 227 228 230 232 233 234 235 236 237 238 239
 1  2  1  2  1  2  2  1  2  4  2  2  4  2  1  1  1  1  1  1  2  1  2  2  2
240 241 242 243 244 245 246 248 251 252 253 254 255 256 257 259 260 261 262 263 264 266 267 268
 1  1  2  1  2  3  1  2  2  3  7  1  3  3  4  3  1  4  4  5  1  2  3  4  4
269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292
 6  3  4  6  2  3  1  2  3  3  4  1  3  2  2  4  3  7  7  4  3  7  4  6  6
293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316
 1  5  3 10  6  3  4  4  5  3  6  3  8  4  8  6  5  9  8 11  5 10  5  6  6
317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
 6 11  6  7  4  4  7  6  6  8  7 10  5 11  9 10  7 11  7  7 11 11 15  9  9
341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364
10  7 16 17 13 15 15 10 13 18 16 15 15 12 13 13 11 10 16 13 13 17 21 10 10
365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388

```

Figure 15.5

[Note – Here also only a part of the output Screen has been shown]

It can be noticed from the above output that the frequency of the numbers are gradually increasing as the same is approaching the Mean (that is 500)

Thereafter the “barplot” function takes this table of frequencies and creates the chart shown above.

Advantages of R

- ⊙ R is the most comprehensive statistical analysis package. As a result new technology and concepts often appear for the first time in R.
- ⊙ As R is an open source language, it can be used anywhere at any point of time.
- ⊙ R Programming is a cross platform language and can be used with any popular operating system.
- ⊙ In R, everyone is welcome to provide new packages, bug fixes and code enhancements.

Disadvantages of R

- ⊙ In the R Programming language, the standard of some packages is less than perfect
- ⊙ Though R commands give little pressure to memory management, but still it may consume all available memory.
- ⊙ In R, basically there is nobody to complain if something doesn't work.
- ⊙ R Programming language is much slower than other programming languages such as Python & MATLAB

• Python

Python is a widely used general purpose, high level programming language. It was created by Guido van Rossum in 1991 and further developed by Python Software Foundation. It was designed with an emphasis on code readability. Its syntax allows programmers to express their concepts in fewer lines of code. It allows the users to work quickly and integrate systems more efficiently,

There are two major Python versions – Python 2 and Python 3. Both are quite different.

Reasons for the increasing popularity of Python

- ⊙ Emphasis on code readability, shorter codes and ease of writing.
- ⊙ Programmers can express logical concepts in fewer lines of code in comparison to that in the languages like C++ or Java.

- ⊙ Python supports multiple programming paradigms like object oriented, imperative and functional programming or procedural.
- ⊙ There exists inbuilt functions for almost all the frequently used concepts.
- ⊙ Philosophy is “Simplicity is the best”

Language features

1. Interpreted

- ⊙ There are no separate compilation and execution steps like C and C++
- ⊙ Directly run the program from the source code
- ⊙ Internally, Python converts the source code into an intermediate form called bytecodes which is then translated into native language of specific computer to run it.
- ⊙ No need to worry about linking and loading with libraries etc.

2. Platform Independent

- ⊙ Python programs can be developed and executed on multiple operating system platforms.
- ⊙ Python can be used on Linux, Windows, Macintosh, Solaris and many more

3. Free and open source

- ⊙ Redistributable

4. High level language

- ⊙ In Python there is no need to take care about low level details such as managing the memory used by the program.

5. Simple

- ⊙ Closer to English language. Hence easy to learn.
- ⊙ More emphasis on the solution to the problems rather than the syntax

6. Embeddable

- ⊙ Python can be used within C and C++ program to give scripting capabilities for the program's user.

7. Robust

- ⊙ Exceptional handling features
- ⊙ Inbuilt memory management techniques

8. Rich library support

- ⊙ The Python standard library is very vast.
- ⊙ Known as the “Batteries included” philosophy of Python. It can help in doing various things involving regular expressions, documentation generation, unit testing, threading, databases, web browsers, CGI, email, XML, HTML, WAV files, cryptography, GUI and many more
- ⊙ Besides the standard library, there are various other high quality libraries such as the Python Imaging Library which is an amazing simple image manipulation library.

Current applications of Python

1. A number of Linux distributions use installers written in Python. For example – in Ubuntu we have installer UBIQUITY written in Python.
2. Python has seen extensive use in the Information Security Industry.
3. Raspberry Pi – single board computer use Python as its principal user programming language.
4. Python is used in Game Development areas also.
5. Python has been successfully embedded in a number of Software products as their scripting language. Some of products/ areas making use of Python are
 - ⊙ GNU Debugger uses Python as a “Pretty Printer” to show complex structures such as C++ Containers.
 - ⊙ Python has also been used in Artificial Intelligence.
 - ⊙ Python is often used for natural language processing tasks.

Beginning with Python Programming

1. Finding an Interpreter:

Before starting to write programs in Python, there is always a need to have an Interpreter to interpret and run the programs. It can be mentioned here that now-a-days presence of online interpreters has made it possible to run programs without installing an Interpreter.

For systems working in *WINDOWS* platform:- There are many interpreters freely available to run Python scripts. One such is IDLE (Integrated Development Environment) that comes bundled with the Python software downloaded from <http://python.org/>

For systems working in *LINUX* platform:- Python comes preinstalled with popular Linux distros such as Ubuntu and Fedora. To check which version of Python has been installed, the user has to type “python” in the terminal emulator. The interpreter should start and print the version number.

For systems working in *MAC* platform:- Generally Python 2.7 come bundled with MacOS. Installation of Python 3 has to be done by the user from <http://python.org/>

Following Example of writing a program to print the words “Python Quiz” is considered for illustration.

Example:

After starting the Interpreter, the user has to type the following lines

```
# Script Begins
print("Python Quiz")
# Scripts Ends
```

The Output will be

```
Python Quiz
```

Explanation:

The code written in the first line i.e. # **Script Begins**

In Python, comments begin with #. This statement is ignored by the interpreter & serves as documentation the program code.

The second line is written as **print (“ Python Quiz”)**

To print something in the console, **print ()** function is used.

The third **# Script Ends** is just another comment as in the first line.

Advantages of Python

1. Presence of third party modules.
2. Extensive support libraries (NumPy for numerical calculations, Pandas for Data Analytics etc.)
3. Open source and community development.
4. Versatile, easy to read, learn and write.
5. User friendly data structures.
6. High level language
7. Dynamically typed language (No need to mention the data type. Based on the value assigned, it takes the data type of its own)
8. Object oriented language.
9. Portable and interactive
10. Ideal for prototypes – provide more functionality with less coding
11. Highly efficient (Clean object oriented design of Python provides enhanced process control. The language is equipped with excellent text processing and integration capabilities. It has its own unit testing framework which has made it more efficient)
12. IoT (Internet of Things) opportunities.
13. Interpreted language.
14. Portable across operating systems.

Disadvantages of Python

1. Slow speed of execution compared to C and C++
2. Absence from mobile computing and browsers
3. For the C & C++ Programmers switching to Python can be irritating as the language requires proper indentation of code. Certain variable names commonly used, like SUM are functions in Python. So C and C++ have to look out for these.

V. Tools for Statistical Data Analysis

• SAS (Statistical Analysis System)

SAS is a powerful software that has been the first choice of the private enterprises for their analytics needs for a long time. Its GUI and comprehensive documentation coupled with reliable technical support made it a very good tool for companies.

It is a command driven software developed by the SAS institute. It is mainly used as a tool for all kinds of

statistical modeling, data management and predictive analysis. It can be used to import data from a variety of data sources, analyze and perform analysis on the data and export the outcome in HTML, PDF, Excel, RTF and other formats. SAS also provides a graphical point-and-click user interface for non-technical users. SAS programs have DATA steps, which retrieve and manipulate data, and PROC steps, which analyze the data. Each step consists of a series of statements. We can also write SAS macros which are pieces of code or variables that are coded once and referenced to perform repetitive tasks. In other words, we can write dynamic codes using SAS.

Types of SAS Software

- ⦿ Windows or PC SAS
- ⦿ SAS EG (Enterprise Guide)
- ⦿ SAS EM (Enterprise Miner i.e. for Predictive Analysis)
- ⦿ SAS Stats

Key features of SAS

- ⦿ Easily accessible raw data files & data in from of an external database. Read and write almost any data format!
- ⦿ Manage data using tools for data entry, editing retrieval, formatting & conversion
- ⦿ Analyze data using descriptive, statistics, multivariate techniques, forecasting, modeling, linear programming
- ⦿ Advanced analytics helps you to make changes and improvements in business practices.
- ⦿ Report formation with perfect graphs
- ⦿ Operations research and project Management
- ⦿ Data updating and modification
- ⦿ Powerful data handling language
- ⦿ Excellent data cleansing functions
- ⦿ Interact with multiple host systems

Applications of SAS

SAS is extremely popular in commercial analytics

- **SPSS (Software Package for Statistical Analysis)**

SPSS is another popular Statistical tool. It is used most commonly in the Social Sciences and is considered as easiest to learn among enterprise statistical tools. It is loved by the non-statisticians because it is similar to Excel. As almost all the users of computer are familiar with Excel, they find SPSS very easy to use. But it has the same downside as SAS – it is expensive.

- **EViews**

Statistical and econometric data analysis is done by this tool.

Econometrics deals with three types of data: cross-sectional data, time series data, and panel (longitudinal) data. In a time series you observe the behavior of a single entity over multiple time periods. This can range from high frequency data such as financial data (hours, days); to data observed at somewhat lower (monthly) frequencies, such as industrial production, inflation, and unemployment rates; to quarterly data (GDP) or annual (historical) data. In a cross-section you analyze data from multiple entities at a single point in time.

One big difference between time series and cross-sectional analysis is that the order of the observation numbers does not matter in cross-sections. With time series, you would lose some of the most interesting features of the data if you shuffled the observations. Finally, panel data can be viewed as a combination of time series and cross-sectional data, since multiple entities are observed at multiple time periods. EViews allows you to work with all three types of data.

EViews is most commonly used for time series analysis in academics, business, and government, but you can work with it easily when you have cross-sections and/or panel data. EViews allows you to save results within a program and to “retrieve” these results for further calculations later.

Remember how you calculated confidence intervals in statistics say for a population mean? Basically you needed the sample mean, the standard deviation, and some value from a statistical table. In EViews you can calculate the mean and standard deviation of a sample and then temporarily “store” these. You then work with these numbers in a standard formula for confidence intervals.

While EViews is truly interactive, you can also run a program as a “batch” job, i.e., you write a sequence of commands and then execute the program in one go. In the good old days the equivalent was to submit a “batch” of cards, each containing a single command, to a technician, who would use a card reader to enter these into the computer, and the computer would then execute the sequence of statements. (You stored this batch of cards typically in a filing cabinet, and the deck was referred to as a “file.”) While you will work at first in interactive mode by clicking on buttons, you will very soon discover the advantage of running your regressions in batch mode. This method allows you to see the history of commands, and you can also analyze where exactly things went wrong if there are problems with any of your commands. This tutorial will initially explain the interactive use of EViews, since it is more intuitive. However, we will switch as soon as it makes sense into the batch mode.¹

While EViews produces graphs and charts, these can often be improved upon by saving the data used in these graphs in a spreadsheet or ASCII format, and then to import the data into Excel (or another spreadsheet program you prefer). Even better, since EViews works in a Windows format, it allows you to cut and paste the data into any other Windows-based program.

VI. Industry Specific Tools

There are several Industry specific tools available or built into industry specific ERPs. Such as SAP, Oracle, Microsoft Dynamics, Sage, Tally etc .

3. Visualisation Tools

Visualisation of the results of Data Analytics is the most important part from the point of view of the users. Visuals makes communication about the insights of the data much easy. It makes a lot of sense to see the visual trends, charts, graphs etc. than to go through huge chunks of data tables. Thus in the categorisation of Data Analytics tools there is a special place for Visualisation Tools. These can be taken as Reporting Tools also. In reality there is no such tool in data analytics which is meant only to prepare reports with visuals for the analysis of data done by some other tool. Actually all the BI tools used for Data Analytics have this capability. But of course there is some difference among these tools from the point of view of capability of visual preparation,

Common features of Visualisation Tools can be given as –

- ⊙ Ease of using multiple databases – Visualisation tools should have the option for picking the data from multiple sources like Excel, CSV, SQL Server, My SQL etc. This facilitates picking up of the relevant data without much of transformation.
- ⊙ Capability to build relationships – Capability to build relationship between various datasets is one of the

most important criteria for any Visualisation tool. This makes the data interactive which results in ease of preparing reports using a variety of datasets.

- ⦿ Capability to Attract the users – Attractive depiction of the information using various colours in different types of charts (Pie chart, Bar chart etc.) is one of the prime requirements of visuals. Due to this, the user can pay proper attention to the appropriate information.
- ⦿ Ease of interaction and dynamism – Dynamism in Visualisation tools allow the users to drill information to different dimensions. The charts have become dynamic in Visualisation Tools and the users can interact with the charts as and when he or she hovers over those. Drag and drop of the items of different charts to the respective dashboards can be possible dynamically.
- ⦿ Graphical user interface – The Visualisation Tools should have the option to view the data in different forms like a Geographical Map, Area Chart, Histogram etc. This facilitates the users to get more insights and the data conveys more important messages.

Some of the popular tools with great capability of generating attractive visuals are **Tableau, Power Bi, Zoho Reports** etc. All these are primarily BI Tools with very good capability to generate colourful reports.

EXERCISE**A. Theoretical Questions:****⊙ Multiple Choice Questions**

1. Tableau is a –
 - a. Business Intelligence Tool
 - b. Visualisation Tool
 - c. Both (a) and (b)
 - d. None of the above

2. Which of the following statement is correct?
 - a. Functioning of ETL Tool is same as that of ELT Tool.
 - b. For Data Analytics the purpose of ETL Tool is same as that of ELT Tool.
 - c. Both (a) and (b)
 - d. None of the above

3. Which of the following statement is incorrect?
 - a. Microsoft Excel is most popular among all the available spreadsheets.
 - b. Zoho Analytics is a tool used for Financial Data analysis.
 - c. Visualisation Tools are the Reporting Tools.
 - d. None of the above.

4. Prescriptive Analytics is very important because –
 - a. It tells about the action to be taken.
 - b. It tells about what is likely to happen.
 - c. It tells about how something has happened.
 - d. It tells about what has happened.

5. Which of the following has no relation to Business Intelligence?
 - a. A set of business analytics solutions to retrieve, analyse and transform data into useful business sights
 - b. Visualisation Tools are primarily BI Tools.
 - c. ABS Glue is a tool used for the purpose of Business Intelligence.
 - d. Embedded Analytics is an important part of any Business Intelligence tool.

6. Which of the following is related to Financial Data Analytics?
 - a. Value driver analytics
 - b. Financial ratio analytics

- c. Predictive sales analysis
 - d. All the above
7. Analysis of a dataset has revealed the fact that profit of a business has reduced for the financial year 2023-24. What category of data analytics it comes under?
 - a. Descriptive Analytics
 - b. Predictive Analytics
 - c. Diagnostic Analytics
 - d. Prescriptive Analytics
 8. #Script Ends – is related to which type of programming language?
 - a. R Programming
 - b. SAS
 - c. Python
 - d. SPSS
 9. Which one of the following is a Key feature of SAS language?
 - a. Capability of handling data analysis related to Operations Research and Project Management.
 - b. Capability of report formation with perfect graphs.
 - c. Capability to interact with multiple host systems
 - d. All the above
 10. Which one of the following is not a spreadsheet?
 - a. Google Sheets
 - b. MS Excel
 - c. EViews
 - d. Quip

Answers:

1	2	3	4	5	6	7	8	9	10
c	b	d	a	c	d	a	c	d	c

⊙ **State True or False**

1. Data Analytics encompasses many diverse types of data analysis.
2. Purpose of Predictive Analytics and Prescriptive Analytics are same.
3. There is no hypothesis stated in Confirmatory data analytics.
4. PowerBI is a business intelligence tool but more popular as data visualisation tool.

5. The process of Data Mining is also known as the process of data discovery.
6. Ease of using multiple databases is not a necessity for Visualisation Tools.
7. Use of cloud data storage system to aggregate diverse datasets in a common location is common for Business Intelligence software.
8. Rating of different BI Software products are based on their capability of Data Analytics, Natural language processing, Real time reporting and Embedded analytics
9. In data analytics the term Mart is related to a place where purchase of data is done.
10. R Programming is incapable of handling any statistical data analysis and that's why SAS is developed.

Answer:

1	2	3	4	5	6	7	8	9	10
T	F	F	T	T	F	T	T	F	F

⊙ **Fill in the blanks**

1. EViews is a statistical and _____ tool.
2. Python is closer to _____ language.
3. QuickBooks is a _____ Analytics tool.
4. Tools handling extract, transform and load functions of data are known as _____ tool.
5. CRAN is related to _____

Answer:

1.	Econometric	2.	English
3.	Financial	4.	ETL
5.	R programming		

⊙ **Short essay type questions**

1. Which type of Data Analytics is necessary to know why there is a decline in certain business?
2. “Data Mining is nothing but Data Discovery” – discuss.

⊙ **Essay type questions**

1. What are the different types of Data Analysis Tool? Discuss each briefly.
2. Write short notes on – Exploratory and Confirmatory Data Analytics, Embedded Analytics facility of BI Tools.

References:

- Mishra. S – Financial management and forecasting using business intelligence and big data analytic tools
- www.technologyadvice.com/business-intelligence
- www.pangaeax.com/2022/03/08/5-financial-analytics-tools-for-business
- www.geeksforgeeks.org/r-programming-language-introduction
- www.datamentor.io/r-programming
- www.jotfarm.com/blog/best-spreadsheet-software