DATA SCIENCE

Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions. Data science uses complex machine learning algorithms to build predictive models.

The data used for analysis can be from multiple sources and present in various formats.

Now that you know what data science is, let’s see why data science is essential in the current scenario.

Why is Data science important?

Data science plays an important role in virtually all aspects of business operations and strategies. For example, it provides information about customers that helps companies create stronger marketing campaigns and targeted advertising to increase product sales. It aids in managing financial risks, detecting fraudulent transactions and preventing equipment breakdowns in manufacturing plants and other industrial settings. It helps block cyber-attacks and other security threats in IT systems.

From an operational standpoint, data science initiatives can optimize the management of supply chains, product inventories, distribution networks and customer service. On a more fundamental level, they point the way to increased efficiency and reduced costs. Data science also enables companies to create business plans and strategies that are based on informed analysis of customer behaviour, market trends and competition. Without it, businesses may miss opportunities and make flawed decisions.

Data science is also vital in areas beyond regular business operations. In healthcare, its uses include diagnosis of medical conditions, image analysis, treatment planning and medical research. Academic institutions use data science to monitor student performance and improve their marketing to prospective students. Sports teams analyze player performance and plan game strategies via data science. Government agencies and public policy organizations are also big users.

How Does Data Science Work?

Data science involves a plethora of disciplines and expertise areas to produce a holistic, thorough and refined look into raw data. Data scientists must be skilled in everything from data engineering, math, statistics, advanced computing and visualizations to be able to effectively sift through muddled masses of information and communicate only the most vital bits that will help drive innovation and efficiency.

Data scientists also rely heavily on artificial intelligence, especially its subfields of machine learning and deep learning, to create models and make predictions using algorithms and other techniques.

Life Cycle of Data Science:

1. **Capture**: The gathering of raw structured and unstructured data from all relevant sources via just about any method—from manual entry and web scraping to capturing data from systems and devices in real-time.

2. **Maintain**: This involves putting the raw data into a consistent format for analytics machine learning or deep learning models. This can include everything from cleansing, deduplicating, and reformatting the data, to using ETL (extract, transform, load) or other data integration technologies to combine the data into a data warehouse, data lake, or other unified stores for analysis.

3. **Process**: Here, data scientists examine biases, patterns, ranges, and distributions of values within the data to determine the data's suitability for use with predictive analytics, machine learning, and/or deep learning algorithms (or other analytical methods).

4. **Analyze**: This is where the discovery happens—where data scientists perform statistical analysis, predictive analytics, regression, machine learning and deep learning algorithms, and more to extract insights from the prepared data.

5. **Communicate:** Finally, the insights are presented as reports, charts, and other data visualizations that make the insights—and their impact on the business—easier for decision-makers to understand. A data science programming languages such as R or Python includes components for generating visualizations; alternatively, data scientists can use dedicated visualization tools.

DATA ANALYTICS

Data analysts bridge the gap between data scientists and business analysts. They are provided with the questions that need answering from an organization and then organize and analyze data to find results that align with high-level business strategy. Data analysts are responsible for translating technical analysis to qualitative action items and effectively communicating their findings to diverse stakeholders.

Data Mining is a popular type of data analysis technique to carry out data modelling as well as knowledge discovery that is geared towards predictive purposes. Business Intelligence operations provide various data analysis capabilities that rely on data aggregation as well as focus on the domain expertise of businesses. In Statistical applications, business analytics can be divided into **Exploratory Data Analysis (EDA) and Confirmatory Data Analysis (CDA)**.

EDA focuses on discovering new features in the data and CDA focuses on confirming or falsifying existing hypotheses. Predictive Analytics does forecasting or classification by focusing on statistical or structural models while in text analytics, statistical, linguistic and structural techniques are applied to extract and classify information from textual sources, a species of unstructured data. All these are varieties of data analysis.

What is Data Analytics for Beginners?

Data Analytics refers to the techniques used to analyze data to enhance productivity and business gain. Data is extracted from various sources and is cleaned and categorized to analyze various behavioural patterns. The techniques and the tools used vary according to the organization or individual.

So, in short, if you understand your Business Administration and have the capability to perform Exploratory Data Analysis, to gather the required information, then you are good to go with a career in Data Analytics.

So, now that you know what is Data Analytics, let's quickly cover the top tools used in this field.

Benefits of Data Analytics

The domain of Data Analytics has been embraced by many industries for the outstanding benefits it offers. Data Analytics is a boon to modern-day businesses. Data Analytics helps businesses in making smarter decisions. Data Analytics improves efficiency and controls risks. Data Analytics also results in cost cuttings.

Roles of Data Analyst:

- **Gather Hidden Insights** – Hidden insights from data are gathered and then analyzed for business requirements.
- **Generate Reports** – Reports are generated from the data and are passed on to the respective teams and individuals to deal with further actions for a high rise in business.
- **Perform Market Analysis** – Market Analysis can be performed to understand the strengths and weaknesses of competitors.
- **Improve Business Requirement** – Analysis of Data allows improving Business to customer requirements and experience.

Types of Data Analysis

There are four types of techniques used for Data Analysis:

1. Descriptive Analysis

With the help of descriptive analysis, we analyze and describe the features of the data. It deals with the summarization of information. Descriptive analysis, when coupled with visual analysis provides us with a comprehensive structure of data.

In the descriptive analysis, we deal with the past data to draw conclusions and present our data in the form of dashboards. In businesses, descriptive analysis is used for determining the Key Performance Indicator or KPI to evaluate the performance of the business.

2. Predictive Analysis

With the help of predictive analysis, we determine the future outcome. Based on the analysis of the historical data, we can forecast the future. It makes use of descriptive analysis to generate predictions. With the help of technological advancements and machine learning, we can obtain predictive insights about the future.

Predictive analytics is a complex field that requires a large amount of data, skilled implementation of predictive models and tuning to obtain accurate predictions. This requires a skilled workforce that is well versed in machine learning to develop effective models.

3. Diagnostic Analysis

At times, businesses are required to think critically about the nature of data and understand the descriptive analysis in depth. To find issues in the data, we need to find anomalous patterns that might contribute to the poor performance of our model.

With diagnostic analysis, you can diagnose various problems that are exhibited through your data. Businesses use this technique to reduce their losses and optimize their performances. Some of the examples where businesses use diagnostic analysis are:

- Businesses implement diagnostic analysis to reduce latency in logistics and optimize their production process.
- With the help of diagnostic analysis in the sales domain, one can update the marketing strategies which would otherwise attenuate the total revenue.

4. Prescriptive Analysis

The prescriptive analysis combines insights from all of the above analytical techniques. It is referred to as the final frontier of data analytics. Prescriptive analytics allows companies to make decisions based on them. It makes heavy usage of Artificial Intelligence to facilitate companies into making careful business decisions.

Major industrial players like Facebook, Netflix, Amazon, and Google are using prescriptive analytics to make key business decisions. Furthermore, financial institutions are gradually leveraging the power of this technique to increase their revenue.

Data Analysis Process

1.Business Understanding Whenever any requirement occurs, firstly we need to determine the business objective, assess the situation, determine data mining goals and then produce the project plan as per the requirement. Business objectives are defined in this phase.

2. Data Exploration

For the further process, we need to gather initial data, describe and explore data and lastly verify data quality to ensure it contains the data we require. Data collected from the various sources is described in terms of its application and the need for the project in this phase. This is also known as data exploration. This is necessary to verify the quality of data collected.

3. Data Preparation

From the data collected in the last step, we need to select data as per the need, clean it, construct it to get useful information and then integrate it all. Finally, we need to format the data to get the appropriate data. Data is selected, cleaned, and integrated into the format finalized for the analysis in this phase.

4. Data Modeling

After gathering the data, we perform data modelling on it. For this, we need to select a modelling technique, generate a test design, build a model and assess the model built. The data model is built to analyze relationships between various selected objects in the data. Test cases are built for assessing the model and the model is tested and implemented on the data in this phase.

5. Data Evaluation

Here, we evaluate the results from the last step, review the scope of error, and determine the next steps to perform. We evaluate the results of the test cases and review the scope of errors in this phase.

6. Deployment

We need to plan the deployment, monitoring and maintenance produce a final report and review the project. In this phase, we deploy the results of the analysis. This is also known as reviewing the project.

The complete process is known as the business analytics process.

Tools used in Data Analytics

With the increasing demand for Data Analytics in the market, many tools have emerged with various functionalities for this purpose. Whether open-source or user-friendly, the top tools in the data analytics market are as follows.

- **R programming** – This tool is the leading analytics tool used for statistics and data modelling. R compiles and runs on various platforms such as UNIX, Windows, and Mac OS. It also provides tools to automatically install all packages as per user requirements.
- **Python** – Python is an open-source, object-oriented programming language that is easy to read, write, and maintain. It provides various machine learning and visualization libraries such as Scikit-learn, TensorFlow, Matplotlib, Pandas, Keras, etc. It also can be assembled on any platform like an SQL server, a MongoDB database or JSON

Tableau Public – This is free software that connects to any data source such as Excel,

- corporate Data Warehouse, etc. It then creates visualizations, maps, dashboards etc with real-time updates on the web.
- **QlikView** – This tool offers in-memory data processing with the results delivered to the end-users quickly. It also offers data association and data visualization with data being compressed to almost 10% of its original size.
- **SAS** – A programming language and environment for data manipulation and analytics, this tool is easily accessible and can analyze data from different sources.
- **Microsoft Excel** – This tool is one of the most widely used tools for data analytics. Mostly used for clients' internal data, this tool analyzes the tasks that summarize the data with a preview of pivot tables.
- **RapidMiner** – A powerful, integrated platform that can integrate with any data source types such as Access, Excel, Microsoft SQL, Tera data, Oracle, Sybase etc. This tool is mostly used for predictive analytics, such as data mining, text analytics, machine learning.
- **KNIME** – Konstanz Information Miner (KNIME) is an open-source data analytics platform, which allows you to analyze and model data. With the benefit of visual programming, KNIME provides a platform for reporting and integration through its modular data pipeline concept.
- **OpenRefine** – Also known as GoogleRefine, this data cleaning software will help you clean up data for analysis. It is used for cleaning messy data, the transformation of data and parsing data from websites.
- **Apache Spark** – One of the largest large-scale data processing engines, this tool executes applications in Hadoop clusters 100 times faster in memory and 10 times faster on disk. This tool is also popular for data pipelines and machine learning model development.

DATA SCIENCE vs DATA ANALYTICS

Although the work of data scientists and data analysts are sometimes conflated, these fields are not the same. The term data science analyst just means one or the other.

A data scientist comes in earlier in the game than a data analyst, exploring a massive data set, investigating its potential, identifying trends and insights, and visualizing them for others. A data analyst sees data at a later stage. They report on what it tells them, make prescriptions for better performance based on their analysis, and optimize any data related tools.

The data analyst is likely to be analyzing a specific dataset of structured or numerical data using a given question or questions. A data scientist is more likely to tackle larger masses of both structured and unstructured data. They will also formulate, test, and assess the performance of data questions in the context of an overall strategy.

Data analytics has more to do with placing historical data in context and less to do with predictive modelling and machine learning. Data analysis isn't an open-minded search for the right question; it relies upon having the right questions in place from the start. Furthermore, unlike data scientists, data analysts typically do not create statistical models or train machine learning tools.

Instead, data analysts focus on strategy for businesses, comparing data assets to various organizational hypotheses or plans. Data analysts are also more likely to work with localized data that has already been processed. In contrast, both technical and non-technical data science skills are essential to processing raw data as well as analyzing it. Of course, both roles demand mathematical, analytical, and statistical skills.

Data analysts have less need for a broader business culture approach in their everyday work. Instead, they tend to adopt a more measured, nailed-down focus as they analyze pieces of data. Their scope and purpose will almost certainly be more limited than those of a data scientist.

In summary, a data scientist is more likely to look ahead, predicting or forecasting as they look at data. The relationship between the data analyst and data is retrospective. A data analyst is more likely to focus on specific questions to answer digging into existing data sets that have already been processed for insights.