

BRAINALYST
A Data Driven Company



DATA SCIENCE 360° PROGRAM

2024-25

**Data Science
Trending Courses**



Call: +91 7419915555



Email: training@brainalyst.in

CONTENTS :

- 1. About Us**
- 2. Exploring the Discipline**
- 3. Scope in Market**
- 4. Applications of Data Science**
- 5. Impact of Data Science**
- 6. Tools and Techniques Required**
- 7. Course Curriculum**
 - I. Data Analytics with Excel**
 - II. Data Analytics with VBA for Macros**
 - III. Data Analytics with SQL**
 - IV. Data Analytics & Visualization with Tableau**
 - V. Data Analytics & Visualization with Power BI**
 - VI. Data Science with R**
 - VII. Data Science with Python**
 - VIII. Forecasting and Text Analytics**
- 8. Study Material**
- 9. 360° Student Evaluation**
- 10. Student Life**
- 11. Why Us**
- 12. FAQs**
- 13. Contact**

ABOUT US

At Brainalyst we all focus to work dedicatedly to help the Businesses & Students to master the field of data science. It starts with understanding and working with core technology frameworks that are used to analyze big data.

Brainalyst has a highly skilled team that has an average domain experience of 7+ Years. We Develop courses by consulting the industry leaders as per the current industry requirements. Our absolute qualified team has experience of working with brands like Air India, Anthem Inc., Kotak Mahindra etc. Our main focus is on providing quality education in the field of Data Science. We specialize in Statistical Modeling, Machine Learning, Deep Learning, Artificial Intelligence, Natural Language Processing and Big Data.

Through years of experience we have identified the gaps in the edu-Tech companies. We have designed our programs in such a way that a learner is aware of all the sub-fields of Data Science.



EXPLORING THE DISCIPLINE

The field of Data Science is unique because it is new and multidisciplinary. Data Science is used to make any kind of prediction, forecast, or decision that helps business to run their operations more efficiently.

The sudden revolution in the hardware capabilities of computers and the spread of the internet has led to data being generated at a very high pace. This has caused different sectors of business to store this data as now, the data holds key to solve a lot of business problems which earlier could only be solved by seasoned business professionals. Data Science can be understood as the scientific way of analyzing the data and create predictive models that analyze underlying patterns of the data and establish a relationship between the various variables and the target (something that is to be predicted). Being a relatively new field there is a lot of speculation regarding the difficulty level of this field and has a reputation of being considered as a tough field to break into. There are numerous reasons for data science to be considered hard and while some reasons seem exaggerated, there are aspects of data science that may be considered challenging.

Data :

In the modern age, data can have multiple meanings. However, in essence, data is information. For all practical purposes, it is a numerical information that can be processed by a machine.

Data Science :

When we approach data with a high level of scientific sophistication that involves proper laid down methodologies, standard operating procedures of experimentation, peer reviewed research that involves exploration of new ways of approaching the data based problem then this scientific study of Data is called Data Science.

Multi-Disciplinary Nature :

Data Science is a unique field as it is an amalgamation of multiple streams such as Mathematics, Statistics, Programming, Visualization, Business Interpretation, etc. However, before getting onto such aspects of Data Science, what needs to be understood is the idea behind the term "Data Science".



Statistics :

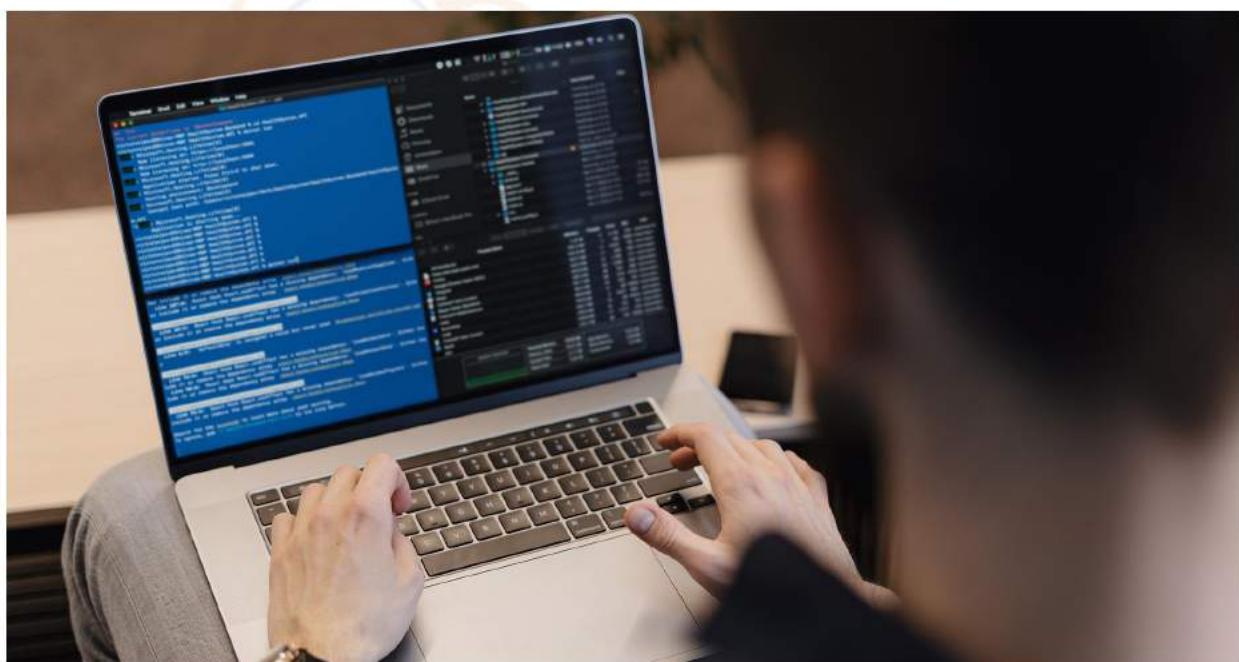
Considered as the backbone of Data Science, Statistics helps the Data Scientists in understanding the underlying patterns present in the data. Through statistics, the user can understand the relationship between various variables that helps in providing a better picture of the given data. Apart from this advanced exploratory data analysis, It also plays an important role in feature engineering which is important for making the data prepared for most algorithms. Additionally, Statistics also acts as a checkpoint for various predictive models and gives insights about the model's inner working,

Mathematics :

Being in the age of Machine Learning and Deep Learning, Mathematics is an important aspect of Data Science. Sophisticated algorithms use advanced concepts of mathematics such as calculus and linear algebra. Having good mathematical knowledge can be considered as a part of data science basics as they provide the data scientist an edge over others in terms of troubleshooting the model's working and tweaking the performance of models using such algorithms.

Programming :

To implement the scientific approaches in order to get meaningful output from data there is a requirement of a tool. While various tools are available out there, most of them require the users to have some basic to intermediate knowledge of programming. While the programming required for implementing data science based concepts may not be as complex as for creating software from the scratch, especially given the modular programming nature of these tools, still, a decent knowledge of programming forms the base of Data Science.



Storytelling (Reporting) and Visualization :

Communication forms another important aspect of Data Science. It may not be as technical as other aspects, but still is a very important and crucial component of any Data Science project. The reason lies in the widespread use of Data Science. Given the amount of data being generated across all sections of business, Data Science is implemented in almost all business domains now which makes it important to report the analysis in simple, easy to understand ways as the people interested in the output may not be Data-oriented.

Thus, providing visual friendly, easily comprehensible, and business logically correct analysis is of paramount importance.



Business Acumen :

Unlike other disciplines, Data Science is not a purely technical field and it requires the data scientist to have a good understanding of various business domains, have problem solving skills and good knowledge of business problems and complexities as only then the insights gained from the implementation of Data Science can be considered useful. In Data Science, it is required not only to analyze the data but to understand it from the prism of business problems and be able to provide a business viable solution.

There is a particular reason because of which Data Science sometimes is considered hard which is the demanding nature of this field.

In order to gain expertise in Data Science, one needs to be good at Mathematics, Statistics, Computer Programming, Visualization, Reporting, Business Understanding, Problem Solving, and Story Telling. As it is an amalgamation of multiple disciplines, this makes it challenging for any individual to master this field as one needs to gain knowledge of all these fields. Aspiring Data Scientists are required to have a knowledge of mathematics and statistics as the numerous predictive algorithms use mathematical and statistical concepts and to troubleshoot a model, these concepts should be known in depth.

The tools of implementation are generally R and Python and they require coding skills.

Data Science can be used anywhere when an informed decision needs to be made. However, such decision making requires various tools and techniques to be implemented at once.



SCOPE IN MARKET

Other fields of studies are quite specialized and specific. Any discipline, for example, Mechanical Engineers, will be required by companies that are involved in this field of study. This in a way restricts the options that a student has in terms of getting placed.

Data Science is a field which is evolving and is high in demand. Data Scientist are required by all the companies all across for a simple reason – All Companies Produce Data.

Data now has the Volume, Variety, Velocity and Veracity that a great deal of information can be extracted from it. This helps companies to take informed decisions. This is the reason that companies look for Data Scientist and hire them if they have the required skillset irrespective of their academic background and other factors.

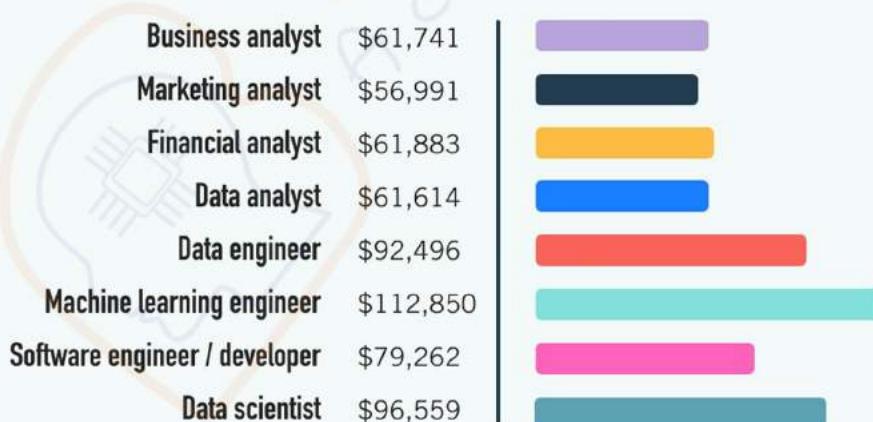
While it will be true to say that Data Science can be applied in any business domain that produces data and as all business domains produce some kind of data, data science can be applied almost everywhere. India Inc is witnessing the rapid digitization of businesses and services, making it the second-largest hub for data science in the world. Analysts predict that the country will have more than 11 million job openings by 2026. In fact, since 2019, hiring in the data science industry has increased by 46%.

Yet, around 93,000 jobs in Data Science were vacant at the end of August 2020 in India. 70% of these vacancies were for positions with less than five years of experience.

In the long term, it would probably be unwise to bet against data science as career move, especially when you widen the field to include related positions like research engineers and machine learning engineers. The U.S. Bureau of Labor Statistics sees strong growth for data science jobs skills in its prediction that the data science field will grow about 28% through 2026.

As technology improves, companies have been able to increase the sophistication of their data operations and analysis. Increasingly, that means inserting artificial intelligence (AI) capabilities into the business processes of regular companies (i.e. non-tech giants). And that means demand for data scientists (average salary in USA \$111,100) and related positions (research scientists and machine learning engineer) will also go up.

Business Analyst Salary vs Other Data Roles



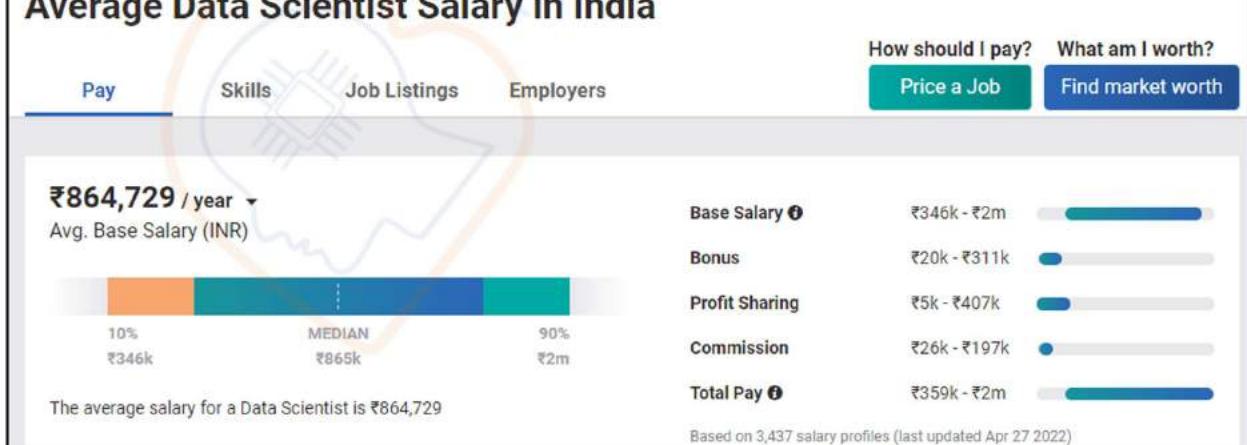
The field of data science and machine learning is quickly becoming one of the most in-demand fields, but there is still a lack of qualified professionals. There are few programs that help but there is a lack of quality and structure in market.

Growth Opportunity and Salary :

If we consider the growth of Data Science based job then it is one of the few fields of study where a fresher can get placed at a 6-digit salary and new joiners often see 75% – 100% hike in their first increment. According to a recent study by LinkedIn, the highest growth jobs are related to data science and data analysis.

| Designation | Growth | Avg Salary |
|----------------------------------|--------|------------|
| Director of Data Science | 9X | 3292635 |
| Data Scientist | 5X | 698413 |
| Full Stack Engineer | 5X | 584992 |
| Machine Learning Engineer | 8X | 565353 |
| Big Data Developer | 5X | 466580 |
| Unity Developer | 1X | 441380 |
| Customer Success Manager | 6X | 425952 |
| Full Stack Developer | 5X | 565353 |
| Sales Development Representative | 7X | 268648 |

Average Data Scientist Salary in India



Data Science situation in India :

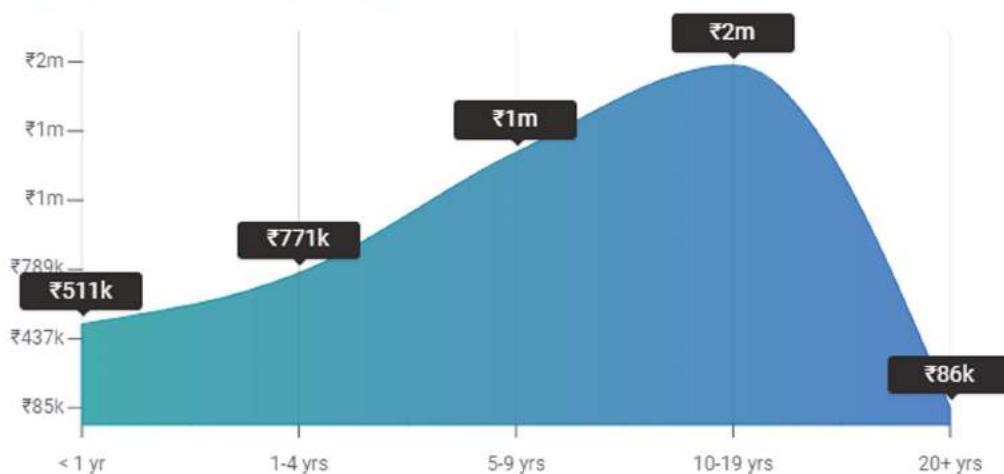
India is witnessing the rapid digitization of businesses and services, making it the second-largest hub for data science in the world. Analysts predict that the country will have more than 11 million job openings by 2026.

Data Science and Machine Learning have a steep learning curve. Even though there is a huge influx of data scientists in India every year, very few people have the required expertise and specialization. As a result, there is a surging demand for professionals with specialized data skills.

Around 93,000 jobs in Data Science are vacant at the end of December 2021 in India. Despite people being laid off in a lot of industries due to covid-19, data science and analytics jobs have an increasing demand not just in India but across the world.

Although the vacant jobs have reduced from 109,000 in February to 82,500 in May 2021, we can see that the demand is consistent for jobs with less than 5 years of experience. Which begs the question, why is there a massive gap in a country that produces over 200,000 Computer Science graduates every year?

The answer to that is simply because Data Science is not a field of IT. IT is a part of Data Science but a Data Scientist a lot of different skill set.



India is witnessing the rapid digitization of businesses and services, making it the second-largest hub for data science in the world.

Analysts predict that the country will have more than 11 million job openings by 2026. Data Science and Machine Learning have a steep learning curve. Even though there is a huge influx of data scientists in India every year, very few people have the required expertise and specialization. As a result, there is a surging demand for professionals with specialized data skills.

List of Top 5 Highest Paying Jobs in India

- ▶ Doctors & Surgeons
- ▶ Data Scientist
- ▶ Investment Banker
- ▶ Software Engineer
- ▶ Commercial Pilots



Scope :

We often use the term 'Scope' in a field. However, Scope is all about the job opportunities in a field and they are determined by three factors- Demand, Supply and Growth.



Lets Understand Them :

Demand :

Data-driven decision making. That is the simple answer to this question. To be a successful company in the 21st century you have to use data to your advantage.

Before many were doing this by using excel to analyze data, but now anyone can have access to and use data-crunching tools like:

- ▶ Google Analytics — Digital marketing cloud-based service
 - Tableau, Power Bi — Data visualization tools for business intelligence
 - ▶ Python, R— Programming languages used to perform complicated analysis with a few lines of code
- The largest companies in the entire world are data science fueled enterprises. Take a look at Google, Amazon, and Facebook. Each use data science to create algorithms that improve customer satisfaction and maximize profits.
- ▶ Google — Ranking of webpages to ensure the top links have an answer to any desired question.
 - ▶ Amazon — Recommendation of products based on consumer's past behavior and interests.
 - ▶ Facebook — Targeted ads (they know the sports you like, preferred price range, food, etc.) to increase market success.

In the end, the main reason demand is still high is because if your competitors are relying on data-driven decision making and you aren't, they will surpass you and steal your market share.

Therefore, companies have to adapt and employ data science tools and techniques or they will simply be forced out of business.

Supply (or the Lack of it) :

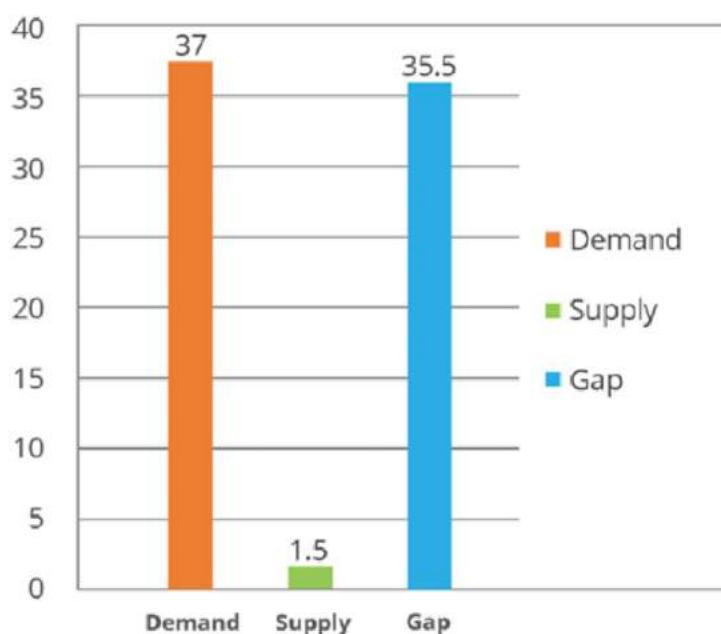
The supply of Data Scientists is low, and it's because the field of data science is still relatively new even in 2021.

You see 20 years ago it was impossible to learn data science because of slow internet connection, and low computational primitive programming languages. As the years went on though, the power of computers started to grow exponentially and data science became possible.

This exponential growth and interest in the field were impossible to predict, and traditional education was not ready to meet the needs of those who wanted to learn this growing field. Very few programs were created to educate aspiring Data Scientists. This shows as research suggests those who get into the field usually transition from other fields such as business, psychology, and life sciences. Most who transitioned learned their skills through self-preparation by reading books, and taking online courses...

Not through the traditional education system.

Lets Understand Them :



Those who get into data science have the advantage of starting a career path in which there are more open jobs than qualified candidates to fill them.

In fact, data science jobs remain open 5 days longer than the average for all other jobs. This points to the fact that there is less competition which results in the recruiters needing extra time to find the correct candidates.

These correct candidates are in luck as most will only need a bachelor's degree to get hired. The low supply has resulted in 61% of data scientist positions be available to those with a bachelor's degree, while only 39% will require a master's degree or a PhD.

Reasons for Supply Deficit :

While there is a large number of business problems that can be solved through Data Science and Data Scientist are paid highly, there still seems to be an everlasting supply gap.

This is due to several reasons-

1. Companies are Facing Real Challenges in Organizing Data :

Throughout the IT boom of the late 1990's and early 2000's the focus was on 'lift and shift' of offline business processes into automated computer systems. Over these two decades, there has been consistent electronic content creation, transactional data generation, and streams of data logs. This means every organization is now sitting on a pile of data that it knows can be of rich value but doesn't know how. This gets reflected in Glassdoor's recently released report which highlights the 50 best jobs in recent times. Unsurprisingly, Data Scientist Jobs is at the top spot for the second year in a row with a score of 4.8/ 5.

2. Shortage of Skilled Resources :

A McKinsey Global Institute study states that the US will face a shortage of about 190,000 data scientists and 1.5 million managers and analysts who can understand and make decisions using Big Data by 2018. The demand is particularly acute in India, where the technologies and tools now exist but not the skilled users. In fact, cofounder and CEO of Fractal Analytics, Srikanth Velamakanni says, "There are two types of talent deficits: Data Scientists, who can perform analytics and Analytics Consultant, who can understand and use data. The talent supply for these job titles, especially Data Scientists is extremely scarce, and the demand is huge."

3. Democratization of Data Scientists :

The need for data scientists is no longer restricted to tech giants.

The Harvard Business Review had reported long back that, "Companies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors." This has finally led even mid to small startups to look towards data sciences. In fact, many smaller firms look to hire entry-level data scientists at decent pay. This works well for both. The scientist finds a significant ground to hone his/ her skills while the organization can afford to pay lesser than what it otherwise would have to.



4. Omnipresence of Jobs :

Since industries from manufacturing to healthcare, IT to banking are leveraging data science in some capacity, there is no dearth of Data Science Jobs for anyone who is interested and is willing to work hard. This is not just limited to industries but also across geographies. So, irrespective of someone's geographical placement or current domain, data science and analytics is open for everyone to pursue.

5. Plethora of Roles :

While Data Science Jobs is an overarching term, within its larger meaning many other sub-roles are available. Roles such as that of a Data Scientist, Data Architect, BI Engineer, Business Analyst, Data Engineer, Database Administrator, Data- and Analytics Manager are in high demand.

6. Long Hiring Times :

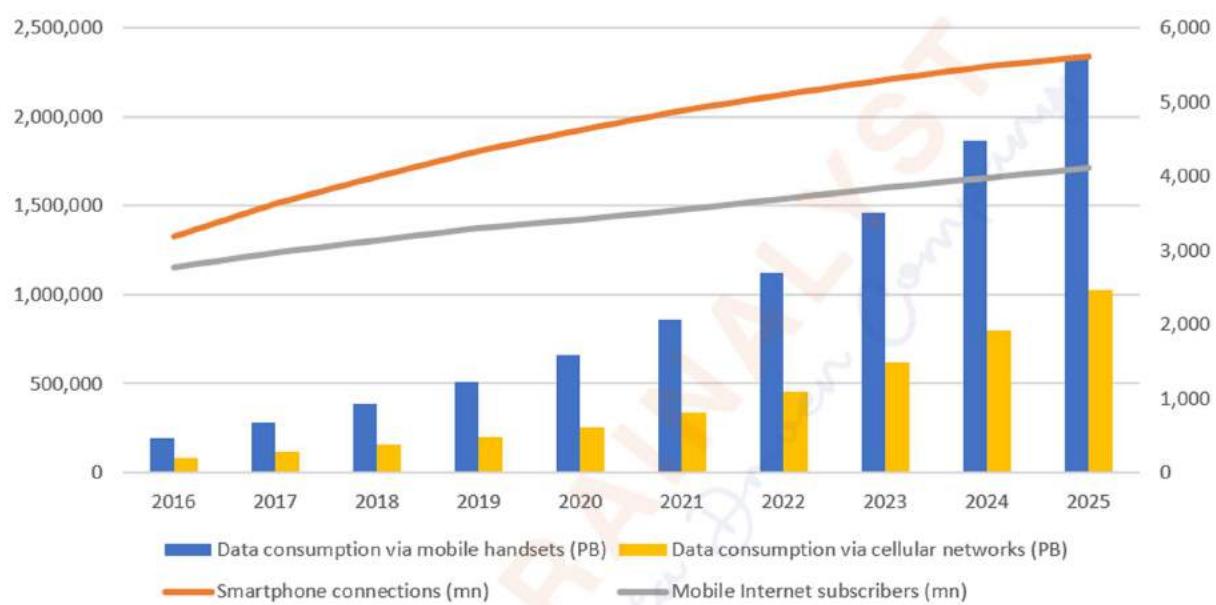
While the time taken to hire engineers is 6-8 weeks, the time to hire data scientists is 11-12 weeks in comparison. The reason behind the vast supply gap and long hiring times can be traced back to existing skill gaps. In fact, since 2019, hiring in the data science industry has increased by 46%.

7. Evolving Tech Trends :

The landscape of business problems that can be solved by Data Science are highly dynamic. Cloud Computing, Computer Vision, Augmented Reality, Blockchain and Quantum Computing are other areas that are being investigated and explored for the near future. All these trends require a data scientist behind them.

8. Availability of Data :

Since internet has become ubiquitous, every action that a user takes online can now be tracked, leading to huge amounts of data. These metrics of human behavior are sorted and analyzed by a data scientist using computer methods. The results then feedback into how a company provides services or sells products.



9. Requirement Across all, even untraditional Domains:

Data scientists are needed in just about every industry. By looking at the demand of different data science categories, we can see three industries jump out: finance, professional services and IT. However, this leaves 41% of the demand in other industries, including health care and retail, marketing, social research etc. This is significant because it shows that data science jobs are integral to across industries.

| DSA Framework Category | Number of Postings in 2015 | Projected 5-Year Growth | Estimated Postings for 2020 | Average Time to Fill (Days) | Average Annual Salary |
|-------------------------------------|----------------------------|-------------------------|-----------------------------|-----------------------------|-----------------------|
| All | 2,352,681 | 15% | 2,716,425 | 45 | \$80,265 |
| Data-Driven Decision Makers | 812,099 | 14% | 922,428 | 48 | \$91,467 |
| Functional Analysts | 770,441 | 17% | 901,743 | 40 | \$69,162 |
| Data Systems Developers | 558,326 | 15% | 641,635 | 50 | \$78,553 |
| Data Analysts | 124,325 | 16% | 143,926 | 38 | \$69,949 |
| Data Scientists & Advanced Analysts | 48,347 | 28% | 61,799 | 46 | \$94,576 |
| Analytics Managers | 39,143 | 15% | 44,894 | 43 | \$105,909 |

Growth :

As Per LinkedIn, there has been a 650% increase in data science jobs since 2012. Glassdoor gives evidence to this claim as they had about 1700 job postings with data science being the primary role in 2016. That number rose to 4500 in 2018, and sort of flattened out in 2020 at around 6500.

COVID-19 was the big story in 2020, and presumably, the reason for this flattening out. Overall though tech jobs have proven to be resilient during the pandemic, which is now in its tenth month.

In the long term, it would probably be unwise to bet against data science as a career move, especially when you widen the field to include related positions like research engineers and machine learning engineers.



APPLICATIONS OF DATA SCIENCE :

The role of Data Science Applications hasn't evolved overnight. Thanks to faster computing and cheaper storage, we can now predict outcomes in minutes, which could take several human hours to process. Data scientists tackle questions about the future. They start with big data, characterized by the four V's: volume, variety, veracity and velocity. Then, they use it as fodder for algorithms and models. The most cutting-edge data scientists, working in machine learning and AI, make models that automatically self-improve, noting and learning from their mistakes.

Data scientists have changed almost every industry. In medicine, their algorithms help predict patient side effects. In sports, their models and metrics have redefined "athletic potential." Data science applications have even tackled traffic, with route-optimizing models that capture typical rush hours and weekend lulls.

While Data Science is a tool that once understood can be applied in thousands of situations, there are few common ones.

1. BFSI Sector :

Considered as the traditional applicants of Data Science, Banking, Financial Services, and Insurance has deployed tools such as SAS to mine their data and find meaningful insights from them. For example, Banking requires Data Scientists to create models to provide information on customers that can be provided with pre-approved loans. To accomplish such tasks, predictive models are to be created that consider multiple variables before taking a decision.

2. Healthcare :

Among the most crucial areas where data science is applied, healthcare has increasingly become dependent on data science. Today, data science can help in reducing the time to diagnose a patient's disease, create reports, suggesting tentative solutions to simple health-related issues, providing 24x7 assistance through AI-enabled chatbox. The biggest revolution that Data Science has made possible is of identifying diseases before they manifest themselves in full force. This helps in saving the lives of humans suffering from diseases that seem nonthreatening on the surface. Data Science also comprises of computer vision that can assist the doctors in quick analysis of medical reports by detecting tumors, inflammation, and other anomalies.



3. Customer Support :

Providing quick and reliable assistance to its customers has remained a prime objective of all businesses. The major issues faced during customer support are the cost of maintaining a customer assistance team and maintaining uniformity in their responses along with making sure that correct information is passed to the customers. This issue has been greatly addressed through the use of Data Science that enables for more informative auto-email replies that analyses the text and provide appropriate responses. Online virtual assistants in the form of chatboxes have picked up popularity in recent times as they have proven their reliability over time and again.

4. Marketing :

Identifying the correct market, customers, or products to cross-sell is among the most important aspect of Marketing. Data Science through the use of classification and segmentation models allows such businesses to identify potential customers that might be interested in certain kinds of products. This reduces the wastage of resources such as calling customers who will have no interest in a product given their historical data.

5. Anomaly Detection :

Anomaly or as commonly known fraud detection has made the digital world lot safer. Data Science helps here in identifying anomalous looking transactions from millions of transactions and automatically informing the concerned people. It is because of such detections that the chances of timely knowing about a fraud preventing further losses have been made increasingly possible.

6. Travel and Hospitality :

This customer-oriented sector heavily deploys data science as given the globalized world and the sheer amount of people traveling worldwide today requires sophisticated methods to keep track of all the plans. Airlines rely on data science for planning their routes, distributing seats on booking, managing flight delays, and identifying customers' travel patterns. Hotels also use Data Science for increasing customer satisfaction by understanding their preferences, providing them with appropriate rooms, maintaining the rescheduling of bookings in a seamless manner.

7. Loyalty Programs :

As the competition increases, so does the chances of customer churn and customers getting engaged with the competitors. This requires the business today to make sure that they retain their valuable customers. This in turn gives birth to loyalty programs that are heavily data-based. By analyzing the data business can decide what benefits to provide to different kind of customers, how to encourage customers to do more business with them, and making sure that the programs are not designed in such a way that they benefit certain customer more and leave out deserving customer making them dissatisfied.

8. Sports :

Among the lesser known application areas, Sports heavily deploys data science for numerous purposes. For example, coaches and team managers rely on data science to identify those underdog players that have the potential but are getting lost due to the availability of large numbers of players worldwide. This helps in bringing new talent to the forefront. Other usage includes tracking how the players perform, analyzing their performance to help them to perform better, and even predicting the performance of a team given the combination of certain players.

9. Security :

Law Enforcement and other security agencies rely on data science to protect their people. Methods of fingerprint matching and facial recognition have become much more efficient and reliable. Security gadgets are now being created that detect motion and can send SOS alerts to the concerned people in case of illegal trespassing.

To find crimes, law enforcement officials often have to go through a large amount of data such as bank records, telephone activities, testimonies in order to detect patterns that can help in solving the case which can also be done through the implementation of Data Science.

10. Mining :

The most recent addition to the growing list of data science users, the mining industry, has started relying on data science to identify the locations where natural resources can be found. The current, nondata science based method is expensive which requires teams to manually dig and test the soil for potential sites for future full-scale mining. Data Science can help the engineers by analyzing historical data, current soil and climate conditions among other variables and provide them with the potential sites, the type of resources and minerals available there along with details on their composition and quantity. This can help the industry in reducing the cost.

11. Logistical Problems :

Data Science helps in tracking and measuring the KPIs like cost, value, services, and waste at regular intervals will help in preventing losses and disasters and taking corrective actions. Route optimization is the process of determining the shortest possible route to reach a location. Data science can be employed to track the nearest vehicle and information can be shared without delay.

It is necessary to track and predict events and processes that can lead to supply chain disruptions. Data science helps by predicting disruptions, and then alerting that to the respective stakeholders.

12. Internet Search and Recommendation Engine :

From Amazon to Netflix, Google to Goodreads, recommendation engines are one of the most widely used applications of machine learning techniques.

Descriptive data models and collaborative filtering techniques are used to compare one visitor's behaviour to other visitors' journeys and historical data to create a list of relevant recommendations.

Predictive segments help marketers take actions based on variables like likelihood to purchase, customer lifetime value, interest-based segmentation, income prediction and customer life cycle status prediction.

13. Digital Marketing :

By analyzing a digital marketer's spend and acquisition data, a Data scientist can build a spending model which will help utilize the budget better to derive maximum ROI from their allotted budgets.

Data science can be used to determine which channels are giving an adequate return for the digital marketer.

Advanced lead scoring can be done if the digital marketer can accurately segment customers as per their interest, it'll increase the sales department's performance, and ultimately, revenue.

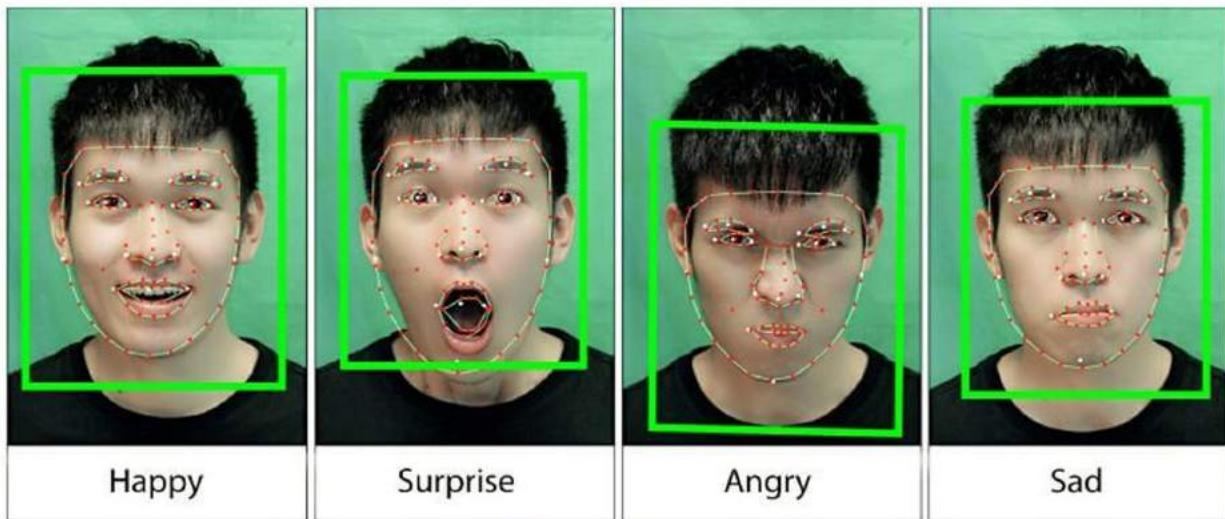


14. Image Processing :

Computer Vision is evolving from the emerging stage and the result is incredibly useful in various applications.

Image Processing provides eyes to the computer that helps in Recognize objects and behaviour, Medical Image Analysis & Diagnosis, Photo Tagging, Face Recognition, Self-Driving Car, Drones etc.

A simple issue such as marking of attendance can be resolved by Data Science.



IMPACT OF DATA SCIENCE :

Data Science has helped and has the potential to further solve those problems that can have a direct impact on the environment in a positive way.

Solutions to unconventional problems (e.g., agriculture) :

Data science can provide actionable insights bespoke to farmers' land and ownership pattern on what to plant, when to plant, and what farm practices to deploy. Making well-informed decisions could save costs and enhance a farmer's profitability and income.

Enabling smart farming. Solutions to enable real-time monitoring of farms will help farmers deploy corrective measures on time.

For example, early disease detection in crops and monitoring soil moisture levels in farms for optimised yield. Smart farming helps farmers monitor relevant parameters continuously through IoT-enabled solutions.

These solutions may also help farmers automate and optimise processes such as pesticide spraying, irrigating, rejuvenation of soil, and other such activities.



Saving Energy :

With technologies like big data and machine learning, employees can now monitor the temperature and performance of production machinery in real-time, and react immediately to solve problems.

Companies will be able to optimize heating, ventilation and air conditioning (HVAC) usage and analyse patterns to deploy energyefficient lighting systems.

It becomes easier for organizations to monitor smart equipment to identify non-performing ones and fix the issue in an efficient manner.

Data Driven Solutions (in Healthcare and Insurance) :

With advanced technology and such a massive volume of data at their disposal, health insurers would be crazy not to use big data analytics to their advantage—especially when it's the key to solving one of the health insurance industry's biggest challenges.

The reality is that health insurance companies are no longer able to compete on the strength of their health plans alone; today's customer expects total transparency and an exceptional experience at every stage of the member lifecycle. Based on this shift in the market, health insurers need provide more insightful recommendations to members based on their personal data so they can make better decisions regarding their coverage and overall health.

TOOLS & TECHNIQUES REQUIRED:

The field of Data is particularly challenging because of the sheer number of tools and techniques required by an individual to master this field.

At every step of the model building process, a different set of tools and techniques are required. Each of these tools help in various aspects. There are several tools that allow us to perform data science, however, each one of them handles some specific aspect of the scientific process. With each one of these tools helping in solving the various problems of business, one must try to know as many of these tools as possible. A practical approach can be to learn any one tool from Python and R for modeling, choose from Tableau or Power BI for visualization, Apache or Hadoop for dealing with big data and Excel and SQL being the must-know tools as they are used almost in all the business organizations.

TOOLS

- 1. Ms Excel, VBA**
- 2. Tableau, SQL**
- 3. R, Python**
- 4. Hadoop, Hive, Spark**
- 5. Redis, AWS, Airflow**
- 6. Scala, Kafka, Hbase**

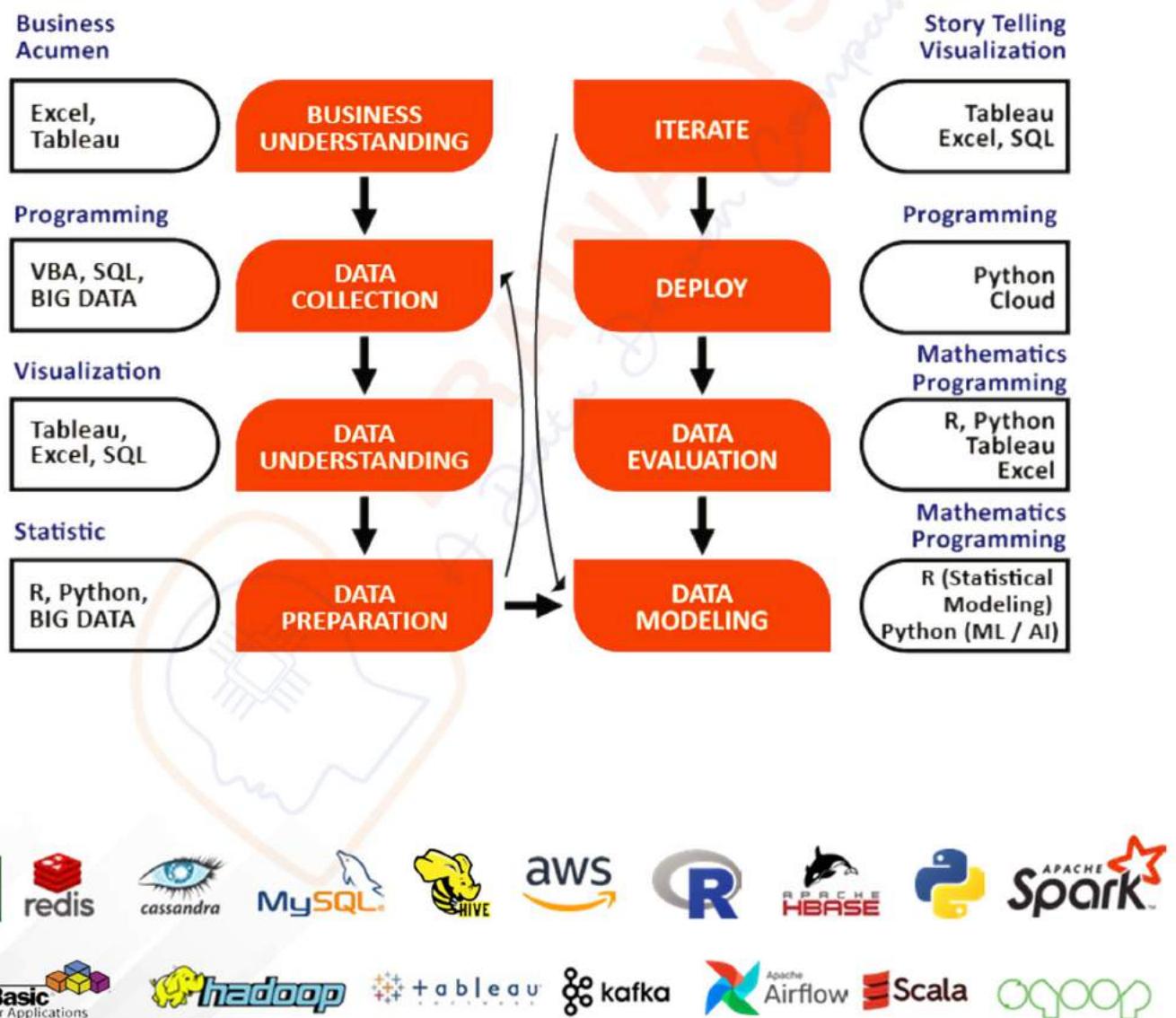
TECHNIQUES

- 1. Mathematics**
- 2. Statistics**
- 3. Data Structures**

Requirement of different techniques and tools at every step of the process:

Different tools aid in different processes. Few helps in data fetching and storage, other helps in reporting and visualization. Another set of tools are required to create and evaluate models along with some basic tools for causal analysis.

For each task, there are several options that one can chose from. But knowing what is required currently in the market should be known beforehand.



Data Science Tools

| Type | Purpose | Common Tools |
|-----------------------------|--|--|
| Collection and Storage | These tools are responsible for storing often large amount of data and help in extracting relevant data in an easy and quick manner | Hadoop, Apache Spark, Apache Hive, Apache Cassandra, Amazon Redshift |
| Analytics | A number of tools allow the user to gain quick insights from the data. There are several APIs that allow for specific analytics with relative ease and higher efficiency | KNIME, Rapid Miner, Splunk, TIBCO Spotfire, Qlik, SQL, MS Excel |
| Reporting and Visualization | One of the most important aspect of analytics is to present often complex information in easy-to-understand format and this is where the reporting aspect of analytics comes into play. Various tools help in creating reports and help especially in the visualization aspect by easing up the process of creating complicated graphs | Tableau, MS Excel, Power BI, Chartio, Redash, Google Data Studio |
| Modeling | The backbone of advanced analytics is predictive modeling and a certain statistical and programming tools allow the user to create such models using libraries that help in developing complex statistical, machine learning and deep learning-based models | Python, R, SAS |



While there are more than a dozen tools that a Data Science aspirant needs to learn to enter this field. Few of the most prominent and sort after tools are as follows:

1. Python :

Python is an open-source, high level, scripting language developed by Guido Van Rossem in 1991. It is recently being considered as a data analysis tool because of the wide range of libraries that the python user community has developed over the last few years which has allowed it to compete with the traditional data and business analytics tools. Following are the important features of this tool:

- ▶ Easy learning curve and can be easily learned by those who are new to the world of programming.
- ▶ It is an open-source and object-oriented language which allows the user to add new functionalities making the tool extremely versatile
- ▶ Python can work with a number of IDE (Integrated Development Environment) and notebooks especially Jupyter notebook which makes for the storing, debugging, and reusing of code extremely easy.
- ▶ It is considered one of the fastest languages, however, it requires a large amount of RAM for it to work efficiently.
- ▶ Python can work well with Big Data platforms and has data mining, manipulation as well as model building capabilities. Packages such as pandas, scipy, and numpy allow this tool to perform any kind of data manipulation whereas sci-kit learn, keras and tensorflow provide the much-needed capability to develop machine learning and deep learning algorithm based predictive models. Other packages include stats and statsmodels that help this tool to perform statistical functions and create statistical

- A primarily advanced tool for modeling, python can be used for reporting and visualization with the help of its packages such as matplotlib, seaborn, altair, etc. helping in automatic generation of reports.

2. R :

R is known as the statistical language made by the statisticians for the statisticians. Like Python, R also requires a bit of coding and programming capabilities from the user for it to function. Developed by Ross Ihaka and Robert Gentleman in 1995 (released in 1997), R is an open-source, statistical language that primarily found its place in the world of academia and research and was later adopted by the BFSI companies. Following are the primary features of R

- R has a bit of steep learning curve (if compared to Python), however, once mastered R provides a lot of capabilities that justified the effort to learn it.
- As R is an open-source language, it has a very strong community with CRAN acting as the quality control organ for this community providing good quality, high capability libraries that cover data analytics needs of almost all domains, from medical to insurance.
- R is commonly used with R studio as the preferred IDE. R Studio has an easy to customize layout and has good code debugging capabilities and even allows for the objects of this object-oriented language to be saved in the physical memory for further use. This makes R studio a much more user-friendly IDE making the task for code reusability much easy.
- The biggest accomplishment of R is the sound statistical capabilities that it possesses which has allowed for its widely accepted use in universities and government organizations

packages that have been developed by CRAN (or in other words: by the developers of R) and not by some third party which makes the results of the statistical tests computed from R much more reliable. The libraries such as caret and h2o.ai provide machine learning and deep learning capabilities whereas dplyr and reshape allow for data manipulation to be done easily in R.

- ▶ Reporting and Visualization is one of the best features of R as it has extremely advanced and sophisticated libraries such as gplot and plotly that can create beautiful and complicated graphs. Also, R allows for the user to easily create dashboards using R shiny which makes it a one of its kind tool and with a simple learning curve.

3. MYSQL :

MYSQL is a query-based language which is one of the most commonly used languages. Before the advent of SAS, R, and Python, the most common tools used for any type of analytics were SPSS and SQL. Today, SPSS sees very limited use, however, SQL has successfully transitioned and adjusted itself to the modern-day requirements of Big Data. It is for the following reason that MYSQL has continued to be in popular use-

- ▶ MYSQL can easily be connected with various software that makes it a very attractive DBMS (Database management system). The recent version of MYSQL even provides a high level of data security and support making it to be adopted by a range of companies.
- ▶ The biggest advantage of SQL is the extremely easy learning curve as a large population of analysts is familiar with it as the SQL queries are mostly in almost plain English.
- ▶ The reason that SQL has survived the wrath of time is because of its speed and it is a high-performance tool that

processing of an extremely large amount of queries and is the reason that a number of E-commerce companies prefer to use MYSQL.

- MYSQL falls in a unique place as technically it is a free, open source tool making it inexpensive with the security of a proprietary software allowing for a wide range of companies, from startup to big multinational to opt for it.

4. Microsoft Excel :

Often ignored and not taken as a serious tool for performing data analytics is MS Excel. The main reason that some sections of data analysts do not consider Excel as an important enough tool is mainly because of its limitation in dealing with large amounts of data. However, one must know that not always there is a need to deal with a large amount of data, and often after passing the data through other tools, for microanalysis, MS Excel is the greatest tool and also is the preferred tool for a preliminary inspection of a sample or a subset of a large dataset. The ease with which Excel is able to perform typical day to day business analytics tasks is the reason that almost all companies on the face of earth deploy it and it is expected from all the analysts that they know at least the basics of it. It is for the following reasons that MS Excel is so famous and a widely accepted tool :

- Excel has an extremely easy learning curve. Because of its graphical user interface, it becomes easy to have a connect with the datasets which is highly important especially for those who are new at dealing with structured data.
- Excel being a commercial tool provides great assistance and detailed material regarding the use of the various excel formulas that allow the user to perform often complex analytical procedures.

- Excel can connect to other DBMS especially SQL servers and with certain plugins, the user can perform data manipulation on even large amounts of data too. The GUI environment when combined with these capabilities makes it a unique and preferred tool for data cleaning and basic aggregation.
- One of the less talked about advantages of Excel is the wide range of plugins available for it which can dramatically increase its capabilities. These tools are often domain-specific and as they are screened through Microsoft, they are of superior quality.
- Lastly, a decent enough aspect of Excel is its visualization capabilities. Most of the typical graphs along with few advanced ones can be created in Excel and is the reason that a lot of companies often after cleaning and aggregating the data use Excel's simple graphs to visualize their analysis.

5. Tableau :

While MS Excel can be used for creating graphs and other advanced tools such as SAS, Python and R can also create sophisticated graphs but still, the place of a dedicated visualization tool cannot be taken by any other kind of tool. There are a number of tools for visualizing data such as D3, Vega, Google Charts, High Charts, etc but one of the widely accepted tools for visualization is Tableau. Tableau has the disadvantage of having a lack of support for higher-level SQL queries and cannot deal with extremely large amounts of data but still is highly popular. We can refer to the following features of Tableau for its widespread use-

- Tableau has the advantage to connect with various data sources such as a number of DBMS, OLAP, and spreadsheets such as MS Excel with which it is especially compatible (especially with the pivot table feature). It can also connect with languages such as R and Python after a few adjustments

- The biggest advantage of Tableau is the extreme ease with which one can use it as there is no programming pre-requisite and people with less background of computer sciences can easily learn it. As several times there are dedicated teams for reporting and visualization, knowing Tableau can provide a much-needed credential in a candidate's profile who is trying to enter the field of analytics.
- Continuing the ease of use, the very interface of Tableau allows for quick reporting and creation of advanced graphs. As Tableau has a graphical user interface, most of the graphics can be created using simple drag and drop functionalities.
- which helps in effortless discoveries of patterns and insights. Lastly, Tableau can be used (like R Shiny) to create dashboards which in this case can be created with much ease, can be made to get updated in real-time, and can be shared with clients through social media.

6. Hadoop Distributed File System (HDFS) :

With the advent of the internet and higher computer processing capability, the amount of data being generated has skyrocketed. In order to handle this large amount of data or commonly known as Big Data, a number of tools have been developed that allow for dealing with the ever-increasing Variety, Volume, and Velocity of data and among them is Hadoop. Hadoop works on the MapReduce technology and allows the user to access and process large amounts of structured as well as unstructured data. Being an open-source tool, it has wide acceptance and is a highly efficient and cost-effective tool to deal with a large amount of data as it can work with a cluster of machines without adding any financial cost to the operations. The following features have led to the widespread acceptance of Hadoop-

- ▶ As mentioned above, Hadoop is an open-source platform that makes it a highly attractive option for dealing with big data. The commercial version of it such as Horton and Cloudera are also available at a reasonable cost that provides troubleshooting support and other assistance.
- ▶ The sheer community of Hadoop is one of its impressive feat. Because Hadoop has been in the world of analytics for some time now and has been adopted by a number of companies, it has led to a vibrant community of users.
- ▶ Another reason why Hadoop is a highly cost-efficient tool is its capability to use community storage which helps companies in reducing their storage expense and as it also allows for pooling of hardware which further brings down the cost of maintaining high computational machines for companies.
- ▶ Like the other tools mentioned in this article, Hadoop too can pride itself for the ease with which it can integrate with other tools. Even though it is developed in Java, Hadoop can easily integrate with languages such as Ruby, Groovy, Perl, and Python. It can also change its processing unit from Map Reduce to other newer processing frameworks such as Apache Spark.

7. Spark :

One of the main competitors of Hadoop MapReduce, Apache Spark is considered as the next generation tool for dealing with analytics when a large amount of data is involved. It is also an open-source data analytics tool that has a big data framework and can integrate with Hadoop making it a highly attractive option for those analytical firms that deal with a large amount of data. Following reasons have led to the widespread popularity of this tool-

- ▶ a. Data using Apache Spark can be processed in real-time. This is particularly advantageous in the field of social media analytics, fraud detection among others where the velocity of data is extremely high.
- ▶ When compared to MapReduce, Apache spark has a relatively easy learning curve and doesn't require much coding for it to function properly. This is the reason that a large number of companies adopt it as they can even train their existing workforce to make them work on Apache Spark.
- ▶ Apache can connect with other languages in order to write the code such as Java, Python, and Scala making it a versatile tool and accessible for people belonging to different programming backgrounds.
- ▶ Apache also allows for numerous algorithms to work with it especially Machine Learning algorithms, SQL queries among others making it to not get stuck as just another big data framework.
- ▶ As it is an open-source tool, it too has wide support and a highly informative community allowing for new Apache users to feel confident.
- ▶ Lastly, the biggest advantage of Spark is its speed which is significantly faster than Hadoop which is because of its RAM intensive framework, however, this comes at the cost of it becoming a memory expensive tool.

Data Analytics with Excel

Excel is a Stepping Stone in a Journey of Data Analytics where you Learn, how to store manipulate and analyze, small well structured data sets.



No. of Hours

18 Hours



Assignment & Case Study

15

Overview :

Data Analytics with Excel :

The advanced Excel training is designed to help you become a skilled Data Analyst and is highly recommended when starting out in the field of Data Science, particularly Data Analytics. By using numerous case studies to ensure hands on experience, the advanced Excel module will teach an individual to make insightful dashboards, and will form the basis to later use VBA macros and SQL. Learning Excel will form the basis for dealing with small structured datasets. their coverage and overall health.



Chapter 1.1 : Basic Excel

- Excel Environment
- Key Terminologies
- Short Cuts
- Key Functionalities
- Copy-paste-paste special
- Formatting & conditional Formatting
- Basic Excel Functions – Types of Functions
- Relational operators
- Data Sorting, Filtering and Data Validation
- Understanding of Name Ranges
- Pivot tables – Charts
- Basics of charts

Chapter 1.2 : Basic Programming Elements

- Overview of programming languages
- Basics of programming elements
- variables, data types, data structures, loops, conditional statements, inputs, outputs, functions etc.
- Understanding key terms
- Client/server
- Database
- Hosting/deployment

Chapter 1.3 : Introduction of Basic Statistics

- Introduction to Statistics
- Measures of central tendencies
- Measures of variance
- Measures of frequency
- Measures of Rank
- Basics of Probability, distributions
- Conditional Probability (Bayes Theorem)

Chapter 1.4 : Overview of Dashboard

- What is dashboard & Excel dashboard
- Adding icons and images to dashboards
- Making dashboards dynamic

Chapter 1.5 : Business dashboard Creations

- Management Dashboard for Sales & Services
- Best practices –
 - Tips and Tricks to enhance dashboard designing

Chapter 1.6 : RDBMS AND SQL (Basic)

- Basic RDBMS Concepts
 - o Introduction to Relational Database management system. Why SQL?
 - o A glance at the tool and its advantages & disadvantages
 - o Understanding Schema, ERDs and Metadata
 - o Introduction to MS SQL Server
 - o What is SQL – A Quick Introduction
 - o Installing MS SQL Server for windows
- Introduction to SQL Server Management Studio
 - o Understanding basic database concepts
 - o Getting started

Chapter 1.7 : Data manipulation using functions

- Descriptive functions
- Logical functions: IF, and, or, not
- Date and Time functions
- Text functions
- Array functions
- Use and application of lookup functions
- Limitations of lookup functions
- Using Index, Match, Offset, reverse vlookup

Chapter 1.8 : Data Visualization in Excel

- Overview of chart types – column/bar charts, line/area , pie, doughnut charts, scatter plots
- How to select right chart for your data
- Creating and customizing advance charts – thermometer charts, waterfall charts, population pyramids

Chapter 1.9 : Introduction to Analytics & Data Science

- What is analytics & Data Science?
- Business Analytics vs. Data Analytics vs. Data Science
- Common Terms in Analytics
- Analytics vs. Data warehousing, OLAP, MIS Reporting
- Types of data (Structured vs. Unstructured vs. Semi Structured)
- Relevance of Analytics in industry and need of the hour
- Critical success drivers
- Overview of analytics tools & their popularity

Chapter 1.10 : Create dashboard in Excel - Using pivot controls

- Concept of pivot cache and its use in creating interactive dash- boards in excel
- Pivot table design elements – concept of slicers and timelines
- Designing sample dashboard using Pivot Controls
- Design principles for including charts in dashboards – do's and dont's

Data Analytics with SQL

When Dealing with Data sets, often the data is stored in Database management SQL- a query base language helps in collecting manipulating and analysing such data base



No. of Hours

18 Hours



Assignment & Case Study

15

Overview :

Data Analytics with SQL :

The purpose of this course is to teach you how to extract and analyze data stored in databases using Structured Query Language (SQL). SQL is the most popular data analysis tool for data analysts and data scientists. Most of the data in the world is stored in databases, and learning SQL will make it possible to access and analyze this data easily. This course is suitable for candidates with various quantitative backgrounds who wish to learn job-oriented analytics and reporting skills using MS-Excel, VBA, MS-Access, and SQL, as well as Tableau.



Chapter 2.1 : Intro to RDBMS & Basic SQL

- Basic RDBMS Concepts
 - o Introduction to Relational Database management system. Why SQL?
 - o A glance at the tool and its advantages and disadvantages
 - o Understanding Schema, ERDs and Metadata
 - o Introduction to MS SQL Server
 - o What is SQL – A Quick Introduction
 - o Installing MS SQL Server for windows
- Introduction to SQL Server Management Studio
 - o Understanding basic database concepts
 - o Getting started

Chapter 2.2 : Data Based Object Creation (DDL commands)

- Creating databases and tables. Understanding data types
- Inserting values into the table
- Altering table properties
- Introduction to Keys and constraints
- Creating, Modifying & Deleting Tables
- Create Table & Create Index statements
- Drop & Truncate statements – Uses & Differences
- DDL Statements with constraints
- Import and Export wizard to get the data in SQL server from excel files or delimited files

Chapter 2.3: Data Manipulation

- Data Manipulation statements
- Insert, Update & Delete statements
- Select statement – Sub setting, Filters, Sorting. Removing Duplicates, grouping and aggregations etc.
- Operators, predicates and built in functions(Top, distinct, Limit)
- Where, Group By, Order by & Having clauses
- SQL Functions – Number, Text, Date, etc.
- SQL Keywords – Top, Distinct, Null, etc.
- SQL Operators – Relational (single valued and multi valued), Logical (and, or, not), Use of wildcard operators and wildcard characters, etc.

Chapter 2.4: Accessing data from multiple tables using Select

- Append and Joins
- Union and Union All – Use & constraints
- Intersect and Except statements
- Table Joins – inner join, left join, right join, full join
- Cross joins/cartesian products, self joins, natural joins etc
- Inline views and sub-queries & it's types
- Optimizing your work
- Update operations with and without joins

Chapter 2.5: Advanced SQL

- Creating table copy and database copy
- Views
- Transactions
- Stored Procedures in SQL
- Crud operations using stored procedures
- Window functions in SQL
- Miscellaneous Topics: Rollup and cube

Data Analytics with VBA For Macros (Self-Paced)

Get your first taste in programming with automating task in excel as often operation performed in Excel can be repetitive and here writing here writing macros can help.



No. of Hours

18 Hours



**Assignment &
Case Study**

10

Overview :

Data Analytics with VBA For Macros :

The Microsoft Excel VBA programming language allows you to automate, analyze, and manipulate data in Excel. Excel VBA is widely used for financial modeling and validation in high finance applications and frequency trading applications. The course introduction of various topics such as Logic, VBA, & Macro Recordings, IDE. Additionally, to multiple case studies used in class sessions, the course includes various assignments and projects for self-study and hands-on experience.



Chapter 3.1 : Data Analytics with VBA

- What is Logic?
- What Is VBA?
- Introduction to Macro Recordings, IDE

Chapter 3.2: A look at some commonly used code snippets

- Explore few code snippets to understand the syntax
- Identify the commonly used codes

Chapter 3.3 : Programming Constructs in VBA

- Control Structures
- Looping Structures
- The With- End with Block

Chapter 3.4 : Functions & Procedures in VBA-Modularizing your programs

- Worksheet & workbook functions
- Automatic Procedures and Events
- Arrays

Chapter 3.5 : Objects & Memory Management in VBA

- The NEW and SET Key words
- Destroying Objects – The Nothing Keyword

Chapter 3.6 : Error Handling

- Explore numerous methodologies for handling errors in VBA
- Learn to read error message

Chapter 3.7 : Controlling accessibility of your code - Access Specifiers

- Understanding the access specifiers to set the accessibility of classes, methods, and other members
- Understand syntax used to facilitate the encapsulation of components

Chapter 3.8 : Code reusability - Adding references and components of your code

- Learning to use references to better deal with objects aiding in code reusability

Chapter 3.9 : How VBA works with Excel

- Working in the Visual Basic Editor
- Introducing the Excel Object Model
- Using the Excel Macro Recorder
- VBA Sub and Function Procedures

Chapter 3.10 : Key Component of programming language

- Essential VBA Language Elements
- Keywords & Syntax
- Programming statements
- Variables & Data types
- Comments
- Operators
- Working with Range Objects

Chapter 3.11 : Communicating with your users

- Simple Dialog Boxes
- User Form Basics
- Using User Form Controls
- Add-ins
- Accessing Your Macros through the User Interface
- Retrieve information through Excel from Access Database using VBA



BRAINYST
Data & Business
Analyst

Data Analytics and Visualization with Tableau

Visualisation is an important aspect of Data analytics as it helps in analytics and presenting the facts visually and an interactive data visualization software like tableau certainly helps.



No. of Hours

18 Hours



Assignment & Case Study

5

Overview :

Data Analytics & Visualization with Tableau

A data visualization is a visual representation of information and data. A data visualization tool employs visual elements such as charts, graphs, and maps to let users view and understand trends, outliers, and patterns in data. As part of this course, you'll be able to learn skills such as MIS Reporting Analyses, Data Mining & Analysis, Data Visualization, and Data Blending & Manipulation. Additionally, to multiple case studies used in class sessions, the course includes various assignments and projects for self-study and hands-on experience.

Chapter 4.1 : Getting Started

- What is Tableau?
- Tableau product suite
- How Does Tableau Work?
- Tableau Architecture
- Connecting to Data & data source concepts
- Understanding the Tableau workspace
- Dimensions and Measures
- Data Types & Default Properties
- Tour of Shelves & Marks Card
- Using Show Me
- Saving and Sharing your work-overview

Chapter 4.2 : Data handlings & Summaries – I

- Date Aggregations and Date parts
- Cross tab & Tabular charts
- Totals & Subtotals
- Bar Charts & Stacked Bars
- Line Graphs with Date & Without Date
- Tree maps/Scatter Plots
- Individual Axes, Blended Axes, Dual Axes & Combination chart/Edit axis
- Parts of Views
- Sorting
- Trend/Reference Lines/Forecasting
- Filters/Context filters

Chapter 4.3 : Data handlings & Summaries – II

- Sets (In/Out Sets/Combined Sets)
- Grouping/Bins/Histograms
- Drilling up/down – drill through
- Hierarchies
- View data
- Actions (across sheets)

Chapter 4.4 : Building Advanced Reports/Maps

- Explain latitude and longitude
- Default location/Edit locations
- Building geographical maps
- Using Map layers

Chapter 4.5 : Calculated Fields

- Aggregate vs. Disaggregate data
- Explain – #Number of Rows
- Basic Functions (String/Date/Numbers etc)
- Usage of Logical conditions

Chapter 4.6 : Table Calculations

- Explain scope and direction
- Percent of Total, Running / Cumulative calculations
- Introduction to LOD (Level of Detail) Expressions
- User applications of Table calculations

Chapter 4.7 : Parameters

- Using Parameters in
 - o Calculated fields
 - o Bins/Reference Lines
 - o Filters/Sets
- Display Options (Dynamic Dimension/Measure Selection)
- Create What-If/ Scenario analysis

Chapter 4.8 : Building Interactive Dashboard

- Combining multiple visualizations into a dashboard (overview)
- Making your worksheet interactive by using actions
- Filter/URL/Highlight
- Complete Interactive Dashboard for Sales & Services

Chapter 4.9 : Building Stories

- Story Points
- Options in Formatting your Visualization
- Working with Labels and Annotations
- Effective Use of Titles and Captions

Chapter 4.10 : Work with data

- Multiple Table Join
- Data Blending
- Difference between joining and blending data, and when we should do each
- Toggle between to Direct Connection and Extracts

Chapter 4.11 : Sharing work with others

- Sharing Workbooks
- Publish to Reader/PDF
- Publish to Tableau Server and sharing on the web

Data Science with R

The language R can help you to perform advanced data analytics and being a statistical language assists you in creating statistical analytical scripts and predictive models.



No. of Hours

70 + 50 Hours



Assignment & Case Study

20

Overview :

Data Science with R :

The comprehensive R training program will provide you with handson skills in Data Science with R. Since recent years, R has gained enormous popularity among Data Science practitioners and it is no surprise that the R language is often referred to as a lingua franca for Data Science. Through various examples and projects, this Data Science R course provides an effective introduction to basic data analytics, statistical prediction modelling, and machine learning.

Chapter 5.1 : Introduction to R

- Introduction to R language
- R for Data Science and Data Analytics
- Introduction to R Editors & IDE's (R Studio)

Chapter 5.2 : Installation of R Software

- Instructions for R & R-Studio Installation
- Differences between R-GUI and R Studio
- Overview of R-Studio & Customize Settings

Chapter 5.3 : Basics of R Programming

- Basics of R-Programming - Syntax Rules
 - o Case Sensitivity
 - o Use of Comments
 - o Object Naming Rules
 - o Operators
 - Assignment
 - Arithmetic
 - Relational
 - Logical
 - Package Reference
- Basic R-Commands & Functions
 - o Shortcuts
 - o Base functions (View, print etc)
- Debugging of code
 - o Reading preliminary errors

Chapter 5.4 : Modular Programming – Packages

- Introduction to Modular Programming & Packages
 - o Difference between System and User libraries
- Important packages in R for Data Science/Data Analytics
 - o Important Packages for data manipulation
 - o Important Packages for statistics
 - o Important Packages for visualization
 - o Important Packages for predictive modelling
- Installing & loading Packages
 - o Different methods of installing packages and pros and cons of them
 - Manual
 - `install.packages()`
 - GUI mechanism
 - Different methods of loading packages and pros and cons of them
 - o `require()`
 - o `library()`
 - o GUI mechanism
- Name Spaces & Managing work environment
 - o Concept of Name Space
 - o Use of changing working directory
 - o Creating .RData data files
 - o Saving particular objects
 - o Saving all objects
 - o Auto Replacement of objects upon loading multiple .RData files
 - o Default Working Directory v/s Personalized Working Directory
 - o Removing objects

Chapter 5.5 : Data Types & Data structures

- R-Data types - strings, factors, numeric, date & time
 - o Different type of numeric based data types and differences between them
 - o Character v/s factor data type
 - o Logical data type
 - o Hierarchy of data types
 - o Dates
 - o Converting to date data type using as.Date()
 - o strp codes
 - o epoch and unix time 1970 concept
 - o Problems with date with year without century
 - o Using lubridate for converting to date data type
- Basic Operations: mathematical – string – date
 - o Concept of class
 - o Generic v/s Class based functions
 - o Exploring basic functions
- R-Data Structures - vectors, matrices, lists, data frames etc.
 - o Homogeneous v/s Heterogeneous data structures
 - o 1D v/s 2D v/s 3D data structures
 - o Vectors
 - Definition, Methods of creation and best practices
 - sub setting of elements
 - Concept of indexing, filtering
 - passing index numbers using where()
 - passing logical input using conditions
 - find and replace
 - o Matrix and List
 - Definition
 - Method of Creation
 - sub setting of elements

- o data.frame and data.table
 - Definition
 - Methods of Creation
 - sub setting of elements and best use practices

Chapter 5.6 : Other Programming Elements

- Control flow & conditional statements
 - o range based loops
 - o condition based loops
 - o making non-vectorised operations vectorised
- User defined functions
 - o Creation numerous UDF
- Concept of apply functions
 - o apply family of functions, difference and best use practices

Chapter 5.7 : Importing & Exporting data

- Importing Data from various sources (Csv, text, excel, xml, etc.)
 - o Use of changing directory
 - o Understanding delimited files
 - o importing files in different format
- Viewing Data objects
 - o Using View() function
 - o Exploring dataset
- Exporting Data to various formats
 - o Exporting dataset in multiple formats

Chapter 5.8 : Understanding of data

- Quick checks of data quality using various R-Functions
- View first and last few rows
 - o Column and Row Names
 - o Number of rows and columns
 - o The frequency, distinct values and proportion of the categories in a text based columns
 - o Arithmetic and Statistical properties of the Numerical columns
- Identifying key problems if there are any
 - o Wrong Data Type of the columns (Solution: Typecasting of the columns)
 - o Columns names not being as per R object naming rules (Solution: Renaming of the columns)
 - o Observations having Duplicates (Solution: Duplicate Removal)
 - o Missing Value in columns (Solution: Missing Value Treatment)
 - o Columns having Outliers (Solution: Outlier Capping)

Chapter 5.9 : Data Preparation/manipulation of data

- Overview of data preparation/Data Manipulation
 - o Basic EDA v/s Structure Based Data Manipulation v/s Content Based Data Manipulation
- Data preparation using single table
 - o Extraction of Columns
 - o Appending Columns
 - o Dropping Columns
 - o Renaming Columns
 - o Rearranging Columns
 - o 3 methods of filtering and pros, cons, best use practices

- o 2 methods of sorting, multi column sorting, base use practices
 - o transposing vs reshaping
 - o wide to long and long to wide format
 - o ID v/s numerical v/s categorical variable and confusions associate with it
 - o Binning
 - o One hot encoding
 - o Label encoding
 - o Aggregation
- Data preparation using multiple tables
- o Merging, theory of joins, one to one, one to many and many to many relationship
 - o appending- rows v/s columns and common mistake associated with appending



Chapter 5.10 : Basic Statistics

- Basic Terms
 - o Samples v/s Population
 - o Statistic v/s Parameter
 - o Need of creating sample
 - o Methods of creating samples
 - o Descriptive v/s Inferential Statistics
- Descriptive Statistics
 - o Measure of Frequency
 - Visualization method
 - Frequency Table
 - o Measure of Central Tendency
 - Mean, Median, Mode
 - Pros and Cons
 - o Measure of Variability
 - Variability /s Central Tendency
 - Range v/s IQR v/s Sum of Absolute Deviation v/s Sum of Squared Deviation v/s Variance v/s Standard Deviation
 - o Measure of Shape
 - Symmetrical v/s Asymmetrical shape
 - Normal Distribution – characteristics etc
 - Different form of skewed distribution with examples
 - Kurtosis- Leptokurtic, Platykurtic, Mesokurtic with examples
 - Unimodal v/s Bimodal v/s Multimodal distributions
- Inferential Statistics
 - o Central Limit Theorem
 - o 3 Sigma Rule
 - o Finding Probability for a raw score
 - o Standard Normal Distribution
 - o Learning to read Z table for +ve and -ve z scores

- o Z Test

- Difference between finding probability for a raw score v/s hypothesis testing
- Standard Error
- Null v/s Alternative Hypothesis
- Significance Level/Alpha Value determination
- Confidence Level
- p-value
- One tailed v/s Two Tailed Hypothesis Testing
- Standard Error: definition and usage

- o t test

- One Sample t-test
 - Signal to Noise ratio formula of t formula
 - Best Case Use
- Dependent t test
 - Understanding Overlapping distributions
 - Understanding paired samples
- Independent t test
 - difference between equal and unequal variance

- o f test

- ANOVA
 - Understanding different ANOVA
 - Understanding One Way ANOVA
 - Relationship between f and p value

- o chi-squared test

- Parametric v/s Non-parametric test
- Expected v/s Observed Table
- Chi Square value as an error term

- Case Study
 - o Identification of correct test
 - o Running multiple test for a single problem to understand the involvement of time
- Others
 - o 6 sigma rule
 - o Industries working at different standard deviation
 - o Point Estimate v/s Confidence Interval

Chapter 5.11 : Data Visualization

- Overview of Data Visualization
 - o Business Driven Graphs
 - o Data Driven Graphs
 - o Univariate v/s Bivariate v/s Multivariate Graphs
 - o Use of different libraries
- Mapping Key graphs for various analysis
 - o Histogram
 - o Boxplot
 - o Line Chart
 - o Line of Best Fit
 - o Scatterplot
 - o Bar Chart
 - o Stacked Bar Chart
 - o Dodged Bar Chart
 - o Facet Bar Chart
 - o Pie Chart
 - o Cross Tab
 - o Word Cloud
 - o Radar Chart
 - o Waffle Chart
 - o Maps
 - o Introduction to Shiny

Chapter 5.12 : Shiny

- Intro to R
- UI-SERVER Connection
- Integration of Input Widgets in UI
- Integration of Input Widgets in Server
- Understanding observe, render and reactive functions
- Exploring different web layout using fluidPage
- Using HTML Codes in Shiny
- Applying themes
- Integration of tables
- Downloading from Shiny app
- Creating an EDA Tool

Chapter 5.13 : Introduction to Modelling

- Evolution of Predictive Models
 - o Stage 1 (Rule Based)
 - o Stage 2 (Statistical Models)
 - o Stage 3 (Machine Learning Models)
 - o Stage 4 (Deep Learning Models)
- Definitions
 - o Model, algorithm, dependent variable, independent variable, estimates/beta's etc.
- Types of Models
 - o Business/Statistical Problem
 - Regression
 - Classification
 - Segmentation
 - Forecasting
 - Optimization

- o Learning Setup
 - Supervised
 - Semi Supervised
 - Un-Supervised
 - Re-Inforcement
- o Business Objective
 - Strategic
 - Objective
- o Nature of Algorithm
 - Statistical
 - Mathematical
 - Rule Based
 - Probabilistic
 - Ensemble
- o Linear v/s Non-Linear
- o Parametric v/s Non-Parametric



Chapter 5.14 : Linear Regression

- Understanding Equation
- Line of Best Fit
- Statistical Method of Finding Unknowns
 - o Manually Calculating beta, constant, SSE, SSR, SST, R2, Adjusted R2, MAPE, RMSE and understanding them visually
- Machine Learning Method of Finding Unknowns
- Running Univariate v/s Bivariate v/s Multivariate Regression
- Understanding output: p-value, t-value, f-value
- Understanding beta's impact on Y
- Impact of Standardized data on beta's magnitude
- Understanding one hold out cross validation, under fitting, over fitting
- Case Study
 - o Pre Modeling
 - Basic EDA
 - Data Quality Checks
 - o Modeling
 - Data Preparation 1
 - Correcting Column Names
 - Cleaning Data for Type Casting
 - Uni-variate Analysis
 - Dropping of rows based on missing value in Y variable
 - Creating KPI and other derived variables
 - Dropping ID variable
 - Dividing data into numerical, categorical and dependent datasets
 - Performing Missing and Outlier Treatment on Numerical datasets
 - Performing Missing (Mode Value) Imputation on Categorical dataset.

- Dividing categorical dataset into ordinal and nominal datasets.
 - Performing label encoding on Ordinal and Dummy Variable Creation on Nominal datasets
 - Combining datasets together
- Data Preparation 3
- Feature Reduction methods
 - RFE
 - Stepwise
 - Lasso
 - VIF
 - Cooks Distance
 - Corr X and Y; Corr X and X
 - Picking important variables
- Data Preparation 4
- Splitting data into train and test
 - Developing model and removing less predictive features
 - Running Cook's Distance and removing observations
 - Implementing model on train and test
- Post Modeling
 - Evaluating Model (train v/s test)
 - R2
 - MAPE
 - RMSE/RMSE
 - Creating simple Calculator in MS Excel using final equation

Chapter 5.15 : Logistic Regression

- Understanding different classification problems
- Linear Regression v/s Logistic Regression
- Logistic Regression Equation
- Imbalanced Class Problems and effect on accuracy calculation
- Threshold (Predicted Class) Based Accuracy Metrics
 - o Confusion Matrix
 - o Sensitivity
 - o Specificity
 - o Precision
 - o F1 Score
 - o Accuracy
 - o False Positive Rate
- Non-threshold based Accuracy Metrics
 - o Concordance
 - o AUC (Area Under the Curve) Score
 - o KS Statistic
 - o Decile Analysis
 - o Lift and Gain Chart
- Identification of threshold
 - o Business Scenario
 - o Ready Made rules based on class proportion
 - o ROC Curve
 - o KS Statistic
- Impact of threshold of Sensitivity and Specificity
- Running Logistic Regression on a Dataset
 - o Calculating accuracy metrics using 0.5 threshold value
 - o Calculating accuracy metrics using different threshold values
 - o Applying model on unseen data
 - o Creating simple Calculator in MS Excel using final equation

Chapter 5.16 : Segmentation (Self-Paced)

- Applications
 - o Customer Segmentation
 - o Market Segmentation
 - Risk Profiling
 - Profiling for Advt.
- Types of Segmentation
 - o Heuristic
 - o Scientific
- Type of Heuristic Segmentation
 - o Value Based Segmentation
 - o Life Stage Segmentation
 - o Loyalty Segmentation
 - o RFM Segmentation
- Type of Scientific Segmentation
 - K-means
 - What is Similarity
 - How this Similarity is found
 - Using Distance metric
 - Using Cosine Similarity
 - What are the things to take care of
 - Standardization of Data
 - Randomness Involved
 - Algorithm: K-means
 - Value of K
 - Centroid
 - Process
 - How to Evaluate the model
 - Profiling: Pros and Cons
 - Size of the Clusters
 - Technical Approach
 - o Silhouette Metric
 - o Pseudo-F value

- Understanding Output
 - Centroid Value
 - Class Labels
- Feature Reduction
 - Correlation between the X vars
 - Principal Component
- DBSCAN
 - Definition
 - Epsilon v/s Min Pts
 - Core v/s Border v/s Noise Point
- Hereditary Clustering
 - AGNES and DIANA
 - Different Links
 - How dendrogram gets created and calculation related to it



Chapter 5.17 : Forecasting (Self-Paced)

- Definition
- Regression v/s Forecasting
- Definition
 - o Cross Sectional Data
 - o Time Series Data
 - o Panel Data
- Different Time Periods
- Types of Forecasting
 - o Weather, Demand, Macro Economics, Trading
 - o Short, Medium, Long Term
 - o Univariate, Bivariate, Multivariate
- Pre-requisites for performing Forecasting
- Important Concepts
 - o Lag
 - o Lead
 - o Differencing
 - o Stationary of Data
 - o Auto-Correlation
 - o Partial-Auto Correlation
 - o Random Walk /White Noise
 - o Trend
 - o Seasonality
 - o Cyclicity
 - o Irregularity

- The difference b/w Model - Forecasting and Predictive Models
 - o Difference in sampling
 - o Difference in Missing Value Imputation
 - Consider the 1 past or 1 future value
 - Using a predictive model (KNN Model)
 - o Outlier Treatment
 - o Additional methods of evaluating the model
 - AIC
 - BIC
- Basic Techniques
 - o MA, WMA
- Intermediate Statistical Techniques
 - o Auto Regressive
 - o smoothing (Holt Winters / ETS Models)
 - Single Exponential
 - Double Exponential
 - Triple Exponential
 - o Decomposition
- Advanced Statistical Techniques
 - o ARIMA Family
 - AR:Auto Regressive Model
 - MA:Moving Average Model
 - ARMA:Auto Regressive Moving Average Model
 - ARIMA:Auto Regressive Integrated Moving Average Model
 - o SARIMA
 - o ARIMAX, SARIMAX
 - o Finding p,d,q values- manual v/s automated
 - o How to use Regression for solving forecasting models?

Chapter 5.18 : Machine Learning (Self-Paced)

- Intro
 - o Definition
 - o Hyper-parameters and identification of them
 - o Bias v/s Variance
 - o K-Fold Cross Validation
 - o Concept of optimization & Optimizers (Gradient Descent/Stochastic Gradient Descent/Mini Batch Gradient Descent)
 - o Understanding loss functions
 - o Over and Under Sampling
- K Nearest Neighbor
 - o Characteristics
 - o Intuition behind KNN
 - o KNN for regression/Classification/Imputation of missing's
 - o Pre-requisites and Assumptions
 - o Hyper-parameters and its effect on complexity of model
 - K
 - Distance Metric
 - o Pros and Cons of KNN
- Naïve Bayes
 - o Traditional v/s Bayesian Stats
 - o Bayesian Theorem
 - o Meaning of Naïve Bayes and Assumption
 - o Important Terms
 - Probability (Class Probability)
 - Conditional Probability
 - Evidence
 - Prior, posterior probabilities
 - o Advantages and Disadvantages

- o Types
 - Bernoulli NB
 - Multinomial NB
 - Gaussian NB
- Decision Trees
 - o Rule / Tree Based Algorithms
 - o Features
 - o Pros and Cons
 - o Assumptions
 - o Splitting Criteria
 - o Regression v/s Classification
 - o Types of decision trees
 - CHAID/CART/C5.0
 - o Hyper-parameters and their effects
 - Splitting criteria of tree
 - MSE
 - F-value
 - Gini
 - Entropy
 - Information Gain
 - Stopping criteria of tree
 - Depth of the Tree
 - Number of Nodes
 - Minimum number of obs in the node
 - Number of features
 - Min Samples in Leaf
- Ensemble Learning
 - o Homogeneous v/s Heterogeneous models
 - o Bagging
 - Process
 - Bootstrap Aggregating
 - Hyper Parameters and their effect on complexity of the model

- o Random Forest
 - Bootstrapping + Random Subspaces
 - Hyper Parameters and their effect on complexity of the model
- o Boosting
 - Ada-Boost
 - Gradient Boosting
 - XGBoosting
- o Sequential v/s Parallel Process
- Support Vector Machine
 - o History of algorithm and its development
 - o concept of margin
 - o Optimization + Constraint problem
 - o Loss Function Explanation
 - o Margin v/s Soft Margin
 - o Kernel Trick / Transformation
 - o Hyper parameters and its effect on complexity of the model
 - constraint
 - gamma
 - o Pros and Cons of SVM
 - o Classifier v/s Regressor SVM
- Regularization
 - o Ridge
 - o Lasso
 - o Elastic Net
 - o Regularization for modeling, feature reduction

Chapter 5.19 : Deep Learning Basics (Self-Paced)

- Introduction to Deep Learning
- Types of Deep Learning Algorithms
- History and Stages of Deep Learning algorithms
- NN Mimicking human brain cell
- Global Features
- Simple Neural Network and the XOR Problem
- Architecture of ANN
 - o Neuron processing
 - o Hidden Layer
 - o Activation Function & types of activation function
 - o Weights and Bias
 - o Optimization Function
 - o Backward Propagation
 - o Feed Forward Network
 - o Epoch/Iteration
 - o Batch Size
 - o Learning Rate
 - o Dense Network/Fully connected network
 - o Loss Function



Chapter 5.20 : Text Mining (Self-Paced)

- Introduction to text functions
- REGEX
 - o Meta Characters
 - o Literals
- Feature Reduction
 - o Basic Techniques
 - Regex
 - Remove the numbers
 - Remove symbols from the data
 - Making the text in lower case
 - Remove the commonly occurring words
 - o Advanced Techniques (NLP based techniques)
 - POS Tagging
 - Stemming
 - Lemmatization
 - Stop Words
 - Business Specific Words
- Conversion of Text to Numerical Data
 - o Count Vectorizer
 - o TF-IDF
 - o Word2Vec
- Web Scrapping
- Sentiment Analysis
- Text Classification - Sentiment Analysis – Intent Analysis
 - o Naïve Bayes
 - o SVM
- Text Segmentation - Topic Mining
 - o K-means
 - o LDA

Data Science with Python

Become a Certified data analyst by learning to import, manipulate & visualize data using the programming language Python.



No. of Hours

120 Hours



**Assignment &
Case Study**

30

Overview :

Data Science with Python :

Using Python as a tool for data science, this course introduces the language to beginners. When you have completed the lab exercises, you will be ready to write your first Python script on your own! The course objectives are to give you an introduction to Python as a language for programming, and to give you a taste of how to work with data in Python. The course will prepare you to apply these skills to business problems and to effectively communicate these understandings to internal and external stakeholders.



Chapter 6.1 : Introduction to Data Science and Python

- History and Evolution of Data Science
- Modern day role of Data Science in Companies
- Components of Data Science
- Role of Python in implementation of Data Science
- Python Notebooks and IDEs

Chapter 6.2 : Data Types and Structures of Core Python

- Fundamental Syntax Rules
- Basic Data Types found in Core Python v/s Pandas
- Type Casting methods and best practices
- Understanding the concept of class
- Exploring associates attributes
- Handling Data Structures found in Core Python

Chapter 6.3 : Role of Modular Programming in Python

- Basic of Programming: Loops and Conditions
- Creating User Defined Functions
- Understanding Modules and Packages
- Installing and Loading Python Packages
- Creating and Using own modules and packages
- Dealing with directories and anaconda environments using yml file

Chapter 6.4 : Vectorized Data Structures: Numpy Array and Pandas Series

- numpy's functions
- Using Numpy's Array
- Using Pandas Series (indexing, functions etc.)
- Comparing Array and Series with traditional python data structure (Difference between non-vectorized and vectorized structure)

Chapter 6.5 : Data Mining (Basic)

- Data Importing and Exporting
- Basic Data Exploration

Chapter 6.6 : Data Mining (Intermediate)

- Extraction of Columns
- Adding of Columns
- Renaming of Columns
- Dropping of Columns
- Rearranging of Columns
- Typecasting of Columns
- Filtering and Sorting of Rows

Chapter 6.7 : Data Mining (Advanced)

- Aggregation
- Appending
- Merging
- Transposing

Chapter 6.8 : Data Quality check

- Missing Value Treatment
- Outlier Capping

Chapter 6.9 : Data Visualization

- Histogram
- Boxplot
- Line Chart
- Scatterplot
- Bar Chart
- Stacked Bar Chart
- Dodged Bar Chart
- Pie Chart
- Cross Tab

Chapter 6.10 : Descriptive Statistics

- Understanding difference in Samples and Population
- Understanding difference in Statistic and Parameter
- Understanding the need of creating samples
- Exploring methods of creating samples
- Understanding difference in Descriptive and Inferential Statistics
- Measure of Frequency
- Measure of Central Tendency
- Measure of Variability
- Measure of Shape

Chapter 6.11 : Understanding Probability Distribution

- Central Limit Theorem
- 3 Sigma Rule
- Finding Probability for a raw score
- Standard Normal Distribution

Chapter 6.12 : Hypothesis Testing

- Z-test
- T-test
- F-test
- Chi-Squared test
- Pearson's Correlation test

Chapter 6.13 : Finding Business Insights using Statistics

- Methods to Identify correct hypothesis test
- Interpreting statistical outputs into business solutions
- Ways to maximise insights gained from hypothesis tests
- 6 sigma rule
- Point Estimate v/s Confidence Interval

Chapter 6.14 : Combining Data Mining and Statistics

- Case Study to explore the approaches used for understanding, manipulating, mining and preparing data

Chapter 6.15 : Introduction to Predictive Modeling

- Converting Business Problems into Data Science based problems
- Evolution of Predictive Modeling Techniques
- Types of Predictive Models
- Types of Learning Algorithms

Chapter 6.16 : Encoding and Binning

- Curse of Dimensionality
- Problem of Multicollinearity
- Types of features
- Binning v/s Encoding
- Label Encoding v/s One-hot Encoding

Chapter 6.17 : Basics of Regression

- Understanding Regression Problems
- Basics of Linear Regression (theory)
- Creating preliminary model using Linear Regression

Chapter 6.18 : Regression Model Building : Pre-Modeling

- Data Quality Checks
- Data Preparation

Chapter 6.19 : Regression Model Building : Modeling

- Assumption Checks
- Feature Reduction
- Data split in development and validation sets

Chapter 6.20 : Regression Model Building - Post Modeling

- Model Evaluation
- Model Interpretation

Chapter 6.21 : Classification Model Building : Pre-Modeling & Modeling

- Data Quality Checks
- Data Preparation
- Assumption Checks
- Feature Reduction
- Data split in development and validation sets

Chapter 6.22 : Classification Model Building : Post Modeling

- Model Evaluation
- Model Interpretation

Chapter 6.23 : Introduction to Machine Learning

- Basics of Machine Learning
- Overfitting - Causes and Solutions
- Resampling
- Regularization

Chapter 6.24 : Decision Trees

- Decision Trees Regressor and Classifier Theory
- Implementing Decision Trees in Python

Chapter 6.25 : K Nearest Neighbour

- KNN Regressor and Classifier Theory
- Implementing KNN in Python



BRAINYST
Data Driven Company

Deep Learning, NLP, Forcasting and Text Mining

Complete your journey in Data science by resolving advanced business problems related to forecasting, clustering, natural language processing and implementation using Python.



No. of Hours

30 Hours



Assignment & Case Study

6

Overview :

Deep Learning, NLP, Forcasting and Text Mining :

We have designed this gold level training course for Data Scientists familiar with Python and interested in creating predictive models that apply Machine Learning to the data in their organization.

Throughout this course you will learn best practices, models, code, algorithms, and a framework for improving your projects. In this course you will learn how to combine state-of-the-art Machine Learning algorithms with Python tools and functionalities to take your Data Science projects to the next level after the completion of the course..

Chapter 7.1 : Introduction to Unsupervised Models

- Understanding Segmentation
- Difference between Supervised and Unsupervised Learning Setup
- Difference between Heuristic and Scientific Segmentation techniques

Chapter 7.2 : Heuristic Segmentation

- Value Based Segmentation
- Loyalty Segmentation
- Life Stage Segmentation
- RFM Segmentation
- Application of Heuristic Segmentation in Python

Chapter 7.3 : Scientific Segmentation : Theory

- K-means
- DBSCAN
- Hierarchical Clustering

Chapter 7.4 : Scientific Segmentation : Application

- Applying Clustering Algorithms in Python
- Exploring methods of evaluation

Chapter 7.5 : Introduction to Forecasting

- Data Preparation for in Forecasting
- Types of Forecasting
- Time Series Components

Chapter 7.6 : Forecasting Algorithms

- Using Averages to Forecast
- Auto Regressive Models
- ETS Models
- Decomposition
- ARIMA/SARIMA
- ARIMAX/SARIMAX

Chapter 7.7 : Solving Univariate and Multivariate Forecasting Problems in Python

- Using forecasting algorithms to solve Univariate Forecasting Problems in Python
- Using forecasting algorithms to solve Bivariate Forecasting Problems in Python

Chapter 7.8 : Introduction to Text Mining

- Understanding the problems with text based data
- Converting unstructured text data into structured data
- Using sklearn to create basic Text Classification model

Chapter 7.9 : Exploring methods of Data Acquisition and Cleaning

- Web Scraping
- Learning Regex

Chapter 7.10 : Feature Reduction in Text Mining using NLP

- POS Tagging
- Stemming
- Lemmatization
- StopWords

Chapter 7.11 : Creating a Text Classification Model

- Performing Sentiment Analysis in Python

Chapter 7.12 : Creating a Text based Segmentation Model

- Performing Topic Mining using K-means
- Performing Topic Mining using LDA

Chapter 7.13 : Interactive Dashboard in Python

- Learning Basics of Streamlit
- Integration Visualization techniques with Streamlit

Chapter 7.14 : Automating Report Generation using Streamlit

- Learning Basics of Streamlit
- Integration Visualization techniques with Streamlit

Chapter 7.15 : Deploying a Prediction Model using Streamlit

- Creating a Loan Approval web-app

Chapter 7.16 : Introduction to Artificial Intelligence

- Evolution of Neural Networks
- Parallels between Neurons And Artificial Neuron
- Single Layer Perceptron: the XOR problem
- Global Features
- Understanding Deep Learning Algorithms

Chapter 7.17 : Artificial Neural Networks - I

- Weights
- Bias
- Activation Functions
- Hidden Layers
- Number of Neurons
- Backpropagation

Chapter 7.18 : Artificial Neural Networks - II

- Understanding Keras and Tensorflow
- Applying ANN in Python for solving Regression Problem
- Applying ANN in Python for solving Classification Problem

Chapter 7.19 : Convolutional Neural Networks - I

- Architecture of CNN
- Layers used in CNN
- Bias
- Pooling and Flattening
- Full Connection

Chapter 7.20 : Convolutional Neural Networks - II

- Implementing CNN in python for Image Classification

Chapter 7.21 : Recurrent Neural Networks - I

- Architecture of RNN
- Layers used in CNN
- Types of LSTM

Chapter 7.22 : Recurrent Neural Networks - II

- Implementing RNN in python for Regression

Chapter 7.23 : Recurrent Neural Networks - III

- Implementing RNN in python for Classification

Chapter 7.24 : Self Organising Maps - I

- Introduction to use of Neural Networks in Unsupervised Problems
- Architecture of SOM
- Similarities to other ML algorithms

Chapter 7.25 : Self Organising Maps - II

- Implementing SOM in python for Customer Segmentation

Chapter 7.26 : Boltzmann Machine - I

- Introduction to Energy Based Models
- Architecture of Boltzmann Machine
- Types of Boltzmann Machine

Chapter 7.27 : Boltzmann Machine - II

- Creating Boltzmann Machine Algorithm using Tensorflow

Chapter 7.28 : Boltzmann Machine - III

- Applying Boltzmann Machine Algorithm to solve Predicative Problem

Chapter 7.29 : Auto Encoders - I

- Architecture of Auto Encoders
- Types of Auto Encoders

Chapter 7.30 : Auto Encoders - II

- Applying Auto Encoders in Python to solve unsupervised learning problem

STUDY MATERIAL :

Brainalyst prides itself in providing quality content its students. Often Data Science aspirant find it difficult to attain academic knowledge regarding the various fields of Data Science and have to buy books and other material from various sources. Also, in order to get ready for the interview, relevant capstone projects and assignments are required.

Brainalyst solves all such problems by providing in-depth study material.

Dedicated Books

For every module, we provide books that are designed as per the curriculum followed in the Data Science 360. These books are in-detail and cover all the relevant points. These helps the students in learning, revising and preparing for tests and interviews. Hard Copies are provided subject to order and chargeable separately¹.

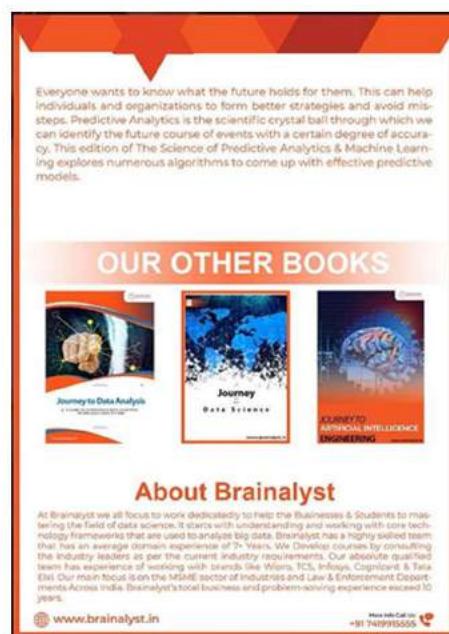
The books have clearly defined chapters so that the learners are informed and can keep themselves up to date throughout the course.

These books cover all the theoretical as well as the practical implementation of the numerous tool and techniques covered in the Data Science Program.

The learner gets the following

- The Stepping Stone in Data Analytics (Excel)
- The Applications of Query Based Language (SQL)
- The Gentle Introduction to Programming (VBA)
- The Tale Explained through Graphs (Tableau)
- The Journey to Statistical Modeling and Beyond (R)
- The Art of Data Analytics (Python-I)
- The Science of Predictive Analytics and Machine Learning (Python-II)
- The Deep End of Data Science (Python -III)
- The Elements of Artificial Intelligence (Python-IV)
- The World of Distributed Computing in Big Data (Big Data – I)
- The Principles of Spark and Cloud Computing in Big Data Big Data – II)

THE SCIENCE OF PREDICTIVE ANALYTICS & MACHINE LEARNING



BOOKS ARE CHARGABLE ADDITIONALLY

Relevant Case Studies & Assignments

Data Science can properly be understood through capstone projects i.e. Case Studies. For every certification program we provide case studies.

These case studies mimic the problems students will face in your real life as a Data Scientist preparing you for the future. The case studies are based on candidates academic and professional background to test and prepare them better.

Interview Material

The learners also get material for effectively preparing for Interviews related to the field of Data Science. These include material on commonly asked questions, guess estimates, aptitude etc.

Other Material

Brainalyst also provides any other relevant reading material like running notes, important books etc. that can aid in the learning experience.

360° STUDENT EVOLUTION :

Brainalyst evaluates the students extensively so that they become master in all the sub-fields of Data Science. This is done through numerous Weekly Tasks, MCQs and Case Studies.

MCQ Exams

At regular intervals MCQ exams are conducted to test the theoretical knowledge of students. This help the students to deal with the numerous conceptual questions related to the use of algorithms, best practices etc.

Practical Exams

CapStone Projects are provided to the students to test the practical applications of the tools and techniques. Such projects are provided at regular intervals to constantly evaluate the student. The student receives their certificate upon the successful completion of the Final exams upon which he/she is enrolled to the job assistance program.

Weekly Tasks

To keep up with every module, students are provided with weekly tasks to aid in the learning process.

STUDENT LIFE :

The student experience is of utmost importance. We make sure that the knowledge imparted throughout the course is consistent and effective. To maximize the student learning experience, we ensure several things.

Batch Size

The maximum batch strength is of ~50 students so that the learning is not cluttered and each student gets the due attention¹¹.

Online Learning Experience

The full stack course is 100% live online. To make sure that everyone is heard, the students are regularly allowed to unmute themselves to resolve any queries¹².

LMS

As all the sessions are live, we make sure that the students are covered if they miss a session. Each student is provided with a custom dedicated Learning Management System where one can find all their sessions recorded so that you don't miss anything¹³.

Assignment, Case Studies and other materials and notifications are provided through the LMS so that the students are never lost.

Doubt Sessions

Doubt Session is conducted on request where students can resolve their queries maximizing their output.

WHY US :

There are several reasons to choose Brainalyst to learn about Data Science.

Topics Covered

The time taken to become a data scientist depends on the individual's ambition and the level they want to achieve. Still, to have a decent knowledge base, one can learn Data Science in **360** hours of training.

We at Brainalyst don't cut any corners and provide training in all the aspects of Data Science so that the student is able to sit confidently in any interview related to the field of Data Science.

Trainers

Our trainers are with professionals with relevant experience and have considerable experience in the world of Data Science. The Board of Directors of Brainalyst comprises of Data Scientist having experience in all fields of Data Science. This enables quality content to the learners.

Specialization

Brainalyst is a data driven company. We specialize in Data Science. This is something we know best and best is what we deliver.

Referral Program

Brainalyst has an interesting and lucrative referral program to help students.

FAQs

Frequently Asked Questions

General Information

Q: What is the Data Science 360 Course?

A: The Data Science 360 Course is a comprehensive program covering all essential aspects of Data Analytics, Data Visualization, Data Science, and Machine Learning. It is designed for beginners and those looking to transition into roles such as Data Analyst, Data Scientist, Business Analyst, or ML Engineer.

Q: Who is this course intended for?

A: This course is ideal for beginners or professionals looking to transition their career into data-related fields. No previous coding or programming experience is required.

Q: Who is eligible to enroll in this course?

A: The course is open to undergraduates, fresh graduates, working professionals, and individuals with study or career gaps. It is a flexible program tailored to meet the needs of a diverse range of learners.

Q: What are the pre-requisites for this course?

A: There are no pre-requisites. This course is beginner-friendly and does not require any prior coding or programming knowledge.

Mode of Delivery

Q: Is the course available online?

A: No, the Data Science 360 Course is a hybrid program available only in online mode. All our live sessions are fully Interactive.

Q: What does "hybrid" mean in the context of this course?

A: A hybrid program means that learners will have access to both self-paced classes and live online interactive sessions. If you prefer to complete the course faster, you can focus on the self-paced modules and still have the opportunity to attend live sessions for additional support and interaction.

Q: If I miss the live classes, will I get access to recordings?

A: Yes, you will get the recordings of all live sessions in your LMS. It usually takes 4-6 hours for the recordings to be available after the class. Additionally, refresher sessions are scheduled during the course tenure to keep all learners on the same page.

Curriculum and Content

Q: What topics are covered in the course curriculum?

A: The curriculum includes:

1. Advanced Excel for Data Analytics
2. VBA for Macros
3. SQL for Data Analytics
4. Data Analytics & Visualization with Tableau
5. Data Analytics & Visualization with PowerBi
6. Data Science with R (Including Machine Learning)
7. Data Science with Python (Including Machine Learning)

Q: How is the course content delivered?

A: The course includes 300+ hours of recorded classes and 180+ hours of live online classes conducted by industry experts.

Q: When are the live classes scheduled?

A: Live classes are held on weekends (Saturday and Sunday) from 10:00 AM to 01:00 PM (IST).

Q: What about projects and assignments?

A: You will receive case studies and assignments for each module, with over 100+ case studies and assignments provided to improve your learning experience. Upon course completion, you will work on two capstone projects—one general and one industry-specific as per your choice, with guidance from an industry expert mentor.

Q: What is the duration of this course?

A: The standard duration is 4+2 months, but you can complete it faster with the self-paced content.

Q: Is LMS access for a lifetime?

A: Lifetime access means until you land a job, but it can be extended upon request from the learner.

Support and Resources

Q: What kind of support is provided to students?

A: Students receive daily doubt sessions available from 10:00 AM to 10:00 PM (IST), dedicated study materials, assignments for each module, industry-similar case studies, and capstone projects.

Q: Are there any additional resources or bonuses?

A: Yes, additional bonuses include:

1. Full Generative AI Course free after completion.
2. Paid internship upon completing the self-paced course in 100 days
3. Referral bonuses
4. Course completion and internship certificates
5. Access to hackathons, coding competitions, and webinars upon course completion
6. Lifetime access to the LMS

Q: Is this a job guarantee program?

A: No, the course does not offer a job guarantee. However, it provides 100 support to help you secure a job. This includes assistance from mentors to prepare an industry-specific CV, optimize your LinkedIn profile, and support through mock interviews and general aptitude test preparation.

Q: What will be the stipend during the internship?

A: The stipend during the internship will be based on your performance and typically ranges from INR 5000 to 12000.

Q: How long is the internship?

A: The internship duration depends on your performance during the first month and can be extended up to three months.

Q: What starting salary (CTC) can a fresher expect after completing the course?

A: A fresher can expect a starting CTC of INR 4 - 7.5 LPA.

Q: What salary hike can experienced professionals expect after completing the course?

A: Individuals with over two years of experience in any domain can expect a minimum 50% and up to a 200% hike on their current CTC.

Fees and Enrollment

Q: What is the fee for the course?

A: The regular price is Rs. 25,999.00 + 18% GST. However, there is a 3rd Anniversary Offer where the entire course is available at discount. This offer is valid for limited time.

Q: Can I pay the fee in installments?

A: Sorry, there is no part payment option available for this program. The course is already offered at a bottom price.

Q: Why is the offer price different on the website?

A: The offer mentioned is a personalized invitation-based offer. If you are interested, you need to share your details to proceed and you will get the course access at the mentioned price.

Q: What is the process of enrollment?

A: To enroll, you need to share your name, phone number, email, sex, and state/country. You will receive an email from Brainalyst containing your user ID and password for the LMS. After logging in with the given credentials, you will see the course on your dashboard and can make the payment to access the course. The enrollment team will guide you through the process.

Enrolment Website: <https://lms.brainalyst.in>

Course Link: <https://lms.brainalyst.in/courses>

Company Information

Q: Where is your company based?

A: The company is based in Gurugram with the headquarters located in Bahadurgarh. We are in market since 2021.

Career Support

Q: What kind of job support is provided?

A: The course offers 100% job support with hiring partners, including resume building, mock interviews, and general aptitude preparation sessions.

Q: Are there mentorship opportunities?

A: Yes, students will have mentorship sessions with industry leaders twice a month. They will share Industry Insights and guide you to secure a job.

THANK YOU

CONTACT US

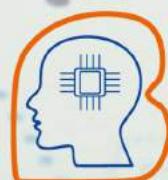
BRAINALYST PVT LTD

Call: +91 7419915555

Email: traning@brainalyst.in

Website: brainalyst.in

A DATA DRIVEN COMPANY



BRAINALYST
A Data Driven Company