

### **1. What is Data Science?**

**Answer:**Data Science is the area of study which involves extracting insights from vast amounts of data using various scientific methods, algorithms, and processes. It helps you to discover hidden patterns from the raw data.

### **2. What is Artificial Intelligence?**

**Answer:**Artificial intelligence (AI) is a set of technologies that enable computers to perform a variety of advanced functions, including the ability to see, understand and translate spoken and written language, analyze data, make recommendations, and more. AI is the backbone of innovation in modern computing, unlocking value for individuals and businesses. For example, optical character recognition (OCR) uses AI to extract text and data from images and documents, turns unstructured content into business-ready structured data, and unlocks valuable insights.

### **3. What is Machine Learning?**

**Answer:**Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

Recommendation engines are a common use case for machine learning. Other popular uses include fraud detection, spam filtering, malware threat detection, business process automation (BPA) and Predictive maintenance.

### **4. Future of Data Science and Real life applications:**

**Answer:**

The future of data science is bright and full of potential. Here are some of the ways that data science will continue to impact real-life applications:

#### **Healthcare**

The healthcare industry, in particular, benefits greatly from data science applications. Data science is making huge strides in the healthcare business. Data science is used in a variety of sectors in health care.

- Image Analysis in Medicine
- Genetics and Genomics
- Drug Development
- Virtual Assistants and Health bots

#### **Drug Development**

From the first screening of medicinal compounds through the prediction of the success rate based on biological variables, data science applications and machine learning algorithms simplify and shorten this process, bringing a new viewpoint to each stage.

Instead of "lab tests," these algorithms can predict how the chemical will behave in the body using extensive mathematical modelling and simulations. The goal of computational drug discovery is to

construct computer model simulations in the form of a physiologically appropriate network, which makes it easier to anticipate future events with high accuracy.

## **E-Commerce**

The e-commerce sector benefits greatly from data science techniques and machine learning ideas such as natural language processing (NLP) and recommendation systems. Such approaches may be used by e-commerce platforms to analyse consumer purchases and comments in order to gain valuable information for their company development.

They utilise natural language processing (NLP) to examine texts and online questionnaires. To evaluate data and deliver better services to its consumers, it is utilised in collaborative and content-based filtering.

## **Text and Advanced Image Recognition**

Speech and picture recognition are ruled by data science algorithms. In our daily lives, we can see the wonderful work of these algorithms. Have you ever needed the help of a virtual speech assistant like Google Assistant, Alexa, or Siri?

Its speech recognition technology, on the other hand, is working behind the scenes, attempting to comprehend and evaluate your words and delivering useful results from your use. Image recognition may be found on Facebook, Instagram, and Twitter, among other social media platforms. When you post a photo of yourself with someone on your profile, these applications offer to identify them and tag them.

## **Gaming**

Machine learning algorithms are increasingly used to create games that grow and upgrade as the player progresses through the levels. In motion gaming, your opponent (computer) also studies your past actions and adjusts its game appropriately. EA Sports, Zynga, Sony, Nintendo, and Activision-Blizzard have all used data science to take gaming to the next level.

## **Security**

Data science may be utilized to improve your company's security and secure critical data. Banks, for example, utilize sophisticated machine-learning algorithms to detect fraud based on a user's usual financial activity. Because of the massive amount of data created every day, these algorithms can detect fraud faster and more accurately than people. Even if you don't work at a financial institution, such algorithms can be used to secure confidential material. Learning about data privacy may help your firm avoid misusing or sharing sensitive information from consumers, such as credit card numbers, medical records, Social Security numbers, and contact information.

data science will continue to have a significant impact on many aspects of our lives, from healthcare and transportation to finance and cybersecurity. As data collection continues to increase, the need for data science will only continue to grow, making it a valuable field with many exciting opportunities for the future.

## **5. Why Python so popular? Future of Python:**

**Answer:**

## **Data Science**

According to Forbes, 2.5 quintillion bytes of data are generated per day in the world. This kind of rapid and extensive creation of data has made Data Science become a lucrative field. Data Science involves the application of various Mathematical and Statistical methods along with Algorithms to find patterns and useful insights in large amounts of structured and unstructured data. Python is one of the most widely used programming languages in Data Science. To get a job in Data Science, having Python skills is highly beneficial.

## **Machine Learning**

Machine Learning (ML) is a branch of AI that works towards making systems perform certain tasks and take decisions without the supervision of human beings. ML is a developing field and has already been used in popular sites like Netflix for movie recommendations. Virtual assistants, chatbots, social media apps, etc use ML. According to Forbes, the global Machine Learning market is expected to reach \$20.83B in 2024. Python has been popular in Machine Learning, and most professionals in this field use Python.

## **Web Development**

Python is a preferred language for Web Development. Many developers find it easy to work with Python since it provides web development libraries and frameworks like Django and Flask. Python is used to develop server-side applications. Strong library support makes development easier as compared to other languages. Python is also highly versatile and creates robust code.

## **Automation**

Since Python is platform-independent and highly versatile, it is used to automate different kinds of applications. A lot of time can be saved if Python is used in automation. Python has a lot of tools, packages, and modules to support the automation of applications quickly. Pytest is a test automation framework that helps in writing Python code for automation. It provides easy creation of test scripts.

## **Simplicity**

Python is a simple language and is easy to understand. This is one of the reasons developers like working on it. It is easy for a new developer to learn Python. It is easily readable as compared to other languages. This is why people prefer using it in various fields and applications. Installing Python is also easy, and it doesn't require much effort. It is platform-independent and can be used across different kinds of applications.

## **Strong Library Support**

Python has many useful libraries that make the development work easy. Programmers don't have to write the entire code from scratch, but instead, they can just import a library to perform complex tasks. Python has libraries for Machine Learning, Data Science, Cybersecurity, etc. Some of the

popular libraries are Numpy, Tensorflow, Keras, Scikit-Learn, etc. New and advanced fields like ML find it easy to implement difficult tasks since Python provides strong library support.

### **Active Communities**

Python has a lot of active communities of programmers across the world. People can share their coding problems, and other programmers provide solutions. Such communities offer a lot of support to programmers while they are working on new applications. They can discuss coding challenges with other community members and find efficient ways to deal with them.

### **Excellent Pay and Immense Career Scope**

Python skilled professionals have a plethora of career options available to them. They usually get paid very well, and there is a lot of scope for growth. Since Python is extensively used across many upcoming technologies, there is a big chance for lucrative job opportunities. Currently, there is a big need for Python programmers since so much innovation, and advancement is happening across fields.

### **Future of Python:**

**Answer:**Python is gradually becoming the preferred language of choice for aspiring programmers who want to master it and get a development job. Python is extremely versatile. As previously said, Python is a multifunctional language that may be used for a variety of purposes. Python is popular for web development, data science, web scraping, automation, etc. It is also used in cybersecurity, blockchain, mobile application development, desktop application development, and many other fields. You can use Python for executing basic tasks as well as highly complex applications, such as analyzing enormous quantities of data. Python is one of the most popular programming languages at the moment, and people prefer it because of its simplicity and versatility. Python is a programming language suitable for everybody, from corporate software developers to casual coders. Data science and AI (artificial intelligence) are relatively young topics still in the early stages of development. Python is the top choice here, which means it isn't going away anytime soon, just as these new fields aren't going away. Data science will be drawing the masses to Python in the future. The relevance of data science is increasing day by day, and as a result, Python is also booming quite a lot. The availability of data science packages like Numpy, Pandas, Matplotlib, etc., makes Python the go-to programming language for data scientists.

Python has a substantial user base in the industry for both computation and building data libraries. Python makes it easy to implement machine learning, which aids in the refinement of algorithm-based technology (for example, speech recognition, product suggestion, etc.).

Python is also preferred for processing and mining massive data, which has a lot of use cases in several industries. It is not an exaggeration to claim that data science and Python will have the finest jobs in the future, complementing each other and propelling them to new heights. Artificial intelligence has a lot to do in many fields, including modern medical science. AI can help cancer research and diagnosis of many diseases at an early stage. AI can do a better job than humans in many cases by analyzing various factors about the disease and patients. The use of technologies like machine learning can immediately enhance the capacity of researchers and medical experts to anticipate treatment outcomes more precisely than ever before. There is no unique programming language devoted to AI, but by examining the properties of each language that can handle AI, they may select Python as it is one of the most reliable and popular programming languages for machine learning.

## 6. Why Python for data analytics?

**Answer:** Python is a multi-functional, maximally interpreted programming language with several advantages that are often used to streamline massive, and complex data sets. Python has a number of distinguishing characteristics that make it the best option for data analysis.

See the features below:

### Easy to learn

Python focuses on both simplicity and readability, while also providing a plethora of useful options for data analysts/scientists. As a result, even novices can easily use its relatively simple syntax to create effective solutions for complex scenarios, with just a few lines of code.

### Flexibility

Python's extreme versatility is another powerful attribute that makes it popular among data scientists and analysts. As a result, data models can be created, data sets can be systematized, ML-powered algorithms can be developed, web services can be developed, and data mining can be used to complete various tasks in a short amount of time.

### Huge libraries collection

It has many completely free libraries that are open to the public. That is a key factor that makes Python essential for data analysis and in data science as well. Users who work in the field of data science are probably familiar with names like Pandas, SciPy, StatsModels, and other libraries that are widely used in the data science community. It's worth noting that the libraries are constantly expanding, providing robust solutions.

### Graphics and visualization

Visual information is well known for being much easier to comprehend, work, and recall. Python provides users with a plethora of different visualization options. As a consequence, it is a must-have method for all data science, not just data processing. By developing numerous charts and graphics, as well as web-ready interactive plots, data analysts can make data more available.

### Built-in data analytics tools

Python's built-in analytics tools make it a perfect tool for processing complex data. Python's built-in analytics tools can also easily penetrate patterns, correlate information in extensive sets, and provide better insights, in addition to other critical matrices in evaluating performance.

## 7. Importance of Python in machine learning:

**Answer:**

**It has a huge number of libraries and frameworks:** The Python language comes with many libraries and frameworks that make coding easy. This also saves a significant amount of time. The most popular libraries are NumPy, which is used for scientific calculations; SciPy for more advanced computations; and scikit, for learning data mining and data analysis.

These libraries work alongside powerful frameworks like TensorFlow, CNTK, and Apache Spark. These libraries and frameworks are essential when it comes to machine and deep learning projects.

**Simplicity:** Python code is concise and readable even to new developers, which is beneficial to machine and deep learning projects. Due to its simple syntax, the development of applications with Python is fast when compared to many programming languages. Furthermore, it allows the developer to test algorithms without implementing them.

Readable code is also vital for collaborative coding. Many individuals can work together on a complex project. One can easily find a Python developer for the team, as Python is a familiar platform. Therefore, a new developer can quickly get acquainted with Python's concepts and work on the project instantly.

**The massive online support:** Python is an open-source programming language and enjoys excellent support from many resources and quality documentation worldwide. It also has a large and active community of developers who provide their assistance at any stage of development. Most scientists have adopted Python for Machine Learning and Deep Learning projects, which means most of the brightest minds worldwide, can be found in Python communities.

**Fast development:** Python has a syntax that is easy to understand and friendly. Furthermore, the numerous frameworks and libraries boost software development. By using out-of-box solutions, a lot can be done with a few lines of code. Python is good for developing prototypes, which boosts productivity.

**Flexible integrations:** Python projects can be integrated with other systems coded in different programming languages. This means that it is much easier to blend it with other AI projects written in other languages. Also, since it is extensible and portable, Python can be used to perform cross languages tasks. The adaptability of Python makes it easy for data scientists and developers to train machine learning models.

**Fast code tests:** Python provides a lot of code review and test tools. Developers can quickly check the correctness and quality of the code. AI projects tend to be time-consuming, so a well-structured environment for testing and checking for bugs is needed. Python is the ideal language since it supports these features.

**Performance:** Some developers argue that Python is relatively slow compared to other programming languages. As much as speed is not one of Python's strong suits, it provides the solution known as Cython. It is a superset of Python language designed to achieve code performance the same as C language. Developers can use Cython to code C extensions the same way they code in Python, as its syntax is almost the same. Cython increases the language performance significantly.

**Visualization tools:** Python comes with a wide variety of libraries. Some of these frameworks offer good visualization tools. In AI, Machine learning, and Deep learning, it is important to present data in a human-readable format. Therefore, Python is a perfect choice for implementing this feature. Some libraries like Matplotlib enable data scientists to generate charts, histograms, and plots to represent data and visualization better. Also, the different APIs that Python supports enhance the visualization process.

## **8. Importance of Python in deep learning & NLP:**

**Answer: Importance of Python in deep learning:**

**Large and Powerful Libraries:** Python has a vast collection of libraries and frameworks that make it easy to perform complex deep learning tasks. The most popular deep learning libraries, such as TensorFlow, PyTorch, and Keras, are written in Python. These libraries provide a wide range of pre-built models, functions, and tools that make it easier to develop deep learning models.

**Easy to Learn:** Python is an easy-to-learn language with a simple syntax, making it accessible to beginners. It is also a very popular language with a large community of developers, making it easy to find support and resources.

### **Importance of Python in NLP:**

**Large and Powerful Libraries:** Python has a vast collection of libraries and frameworks that make it easy to perform complex NLP tasks. The most popular NLP libraries, such as NLTK, spaCy, and Gensim, are written in Python. These libraries provide a wide range of pre-built models, functions, and tools that make it easier to develop NLP models.

**Easy to Learn:** Python is an easy-to-learn language with a simple syntax, making it accessible to beginners. It is also a very popular language with a large community of developers, making it easy to find support and resources.

### **9. Supervised Learning with examples**

**Answer:**Supervised Machine Learning is an algorithm that learns from labeled training data to help you predict outcomes for unforeseen data. In Supervised learning, you train the machine using data that is well “labeled.” It means some data is already tagged with correct answers. Let’s see some Supervised learning examples on how you can develop a supervised learning model of this example which help the user to determine the commute time. The first thing you requires to create is a training set. This training set will contain the total commute time and corresponding factors like weather, time, etc. Based on this training set, your machine might see there’s a direct relationship between the amount of rain and time you will take to get home. So, it ascertains that the more it rains, the longer you will be driving to get back to your home. It might also see the connection between the time you leave work and the time you’ll be on the road. The closer you’re to 6 p.m. the longer it takes for you to get home. Your machine may find some of the relationships with your labeled data.

This is the start of your Data Model. It begins to impact how rain impacts the way people drive. It also starts to see that more people travel during a particular time of day.

### **Types of Supervised Machine Learning Algorithms**

Following are the types of Supervised Machine Learning algorithms:

#### **Regression:**

Regression technique predicts a single output value using training data.

Example: You can use regression to predict the house price from training data. The input variables will be locality, size of a house, etc.

Strengths: Outputs always have a probabilistic interpretation, and the algorithm can be regularized to avoid overfitting.

Weaknesses: Logistic regression may underperform when there are multiple or non-linear decision boundaries. This method is not flexible, so it does not capture more complex relationships.

### **Logistic Regression:**

Logistic regression method used to estimate discrete values based on given a set of independent variables. It helps you to predicts the probability of occurrence of an event by fitting data to a logit function. Therefore, it is also known as logistic regression. As it predicts the probability, its output value lies between 0 and 1.

Here are a few types of Regression Algorithms

### **Classification:**

Classification means to group the output inside a class. If the algorithm tries to label input into two distinct classes, it is called binary classification. Selecting between more than two classes is referred to as multiclass classification.

Example: Determining whether or not someone will be a defaulter of the loan.

Strengths: Classification tree perform very well in practice

Weaknesses: Unconstrained, individual trees are prone to overfitting. Here are a few types of Classification Algorithms

### **Naive Bayes Classifiers**

Naive Bayesian model (NBN) is easy to build and very useful for large datasets. This method is composed of direct acyclic graphs with one parent and several children. It assumes independence among child nodes separated from their parent.

### **Decision Trees**

Decisions trees classify instance by sorting them based on the feature value. In this method, each node is the feature of an instance. It should be classified, and every branch represents a value which the node can assume. It is a widely used technique for classification. In this method, classification is a tree which is known as a decision tree. It helps you to estimate real values (cost of purchasing a car, number of calls, total monthly sales, etc.).

### **Support Vector Machine**

Support vector machine (SVM) is a type of learning algorithm developed in 1990. This method is based on results from statistical learning theory introduced by Vap Nik. SVM machines are also closely connected to kernel functions which is a central concept for most of the learning tasks. The kernel framework and SVM are used in a variety of fields. It includes multimedia information retrieval, bioinformatics, and pattern recognition.

## **10. Unsupervised learning with examples:**

**Answer:** Unsupervised Learning is a machine learning technique in which the users do not need to supervise the model. Instead, it allows the model to work on its own to discover patterns and information that was previously undetected. It mainly deals with the unlabelled data.



## **Example of Unsupervised Machine Learning**

Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs. The algorithm is never trained upon the given dataset, which means it does not have any idea about the features of the dataset. The task of the unsupervised learning algorithm is to identify the image features on their own. Unsupervised learning algorithm will perform this task by clustering the image dataset into the groups according to similarities between images.

Unsupervised Learning algorithms:

Below is the list of some popular unsupervised learning algorithms:

- K-means clustering
- KNN (k-nearest neighbors)
- Hierarchical clustering
- Anomaly detection
- Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Apriori algorithm
- Singular value decomposition

## **11. Role and responsibilities of a Database Engineer:**

**Answer:** Database engineers are the managers of the flow of information at a company. They design and maintain databases that provide informative and relevant data to employees throughout an organization. They work as either applications or system software developers and are generally employed by computer systems companies to design and monitor complex databases. They insure a database is fully operational and provides valid and relevant data. They must implement and maintain computer databases.

Database engineers ensure the database is kept available, secure, and stable. They constantly test its data to ensure it is not corrupted and perform regular updates to keep up with ever-changing computer technologies. They integrate new products and software into a working and current system and see that the data produced is accurate and train employees about how to make the best use of its content. Database engineers need a bachelor's degree in computer science, information technology, or related fields.

### **What responsibilities are common for Database Engineer jobs?**

- Recommend and implement solutions for performance monitoring and tuning.
- Review work of less senior team members, provide guidance from technical perspective.
- Analyze problems, anticipate future problem areas, and implement solutions.
- Perform environment setup and configuration, proactive monitoring and maintenance.
- Lead development of database structures that fit into the overall architecture of the system under development.
- Work with architects and lead developers to come up with optimal database design.
- Engage in capacity planning exercises to ensure adequate system resources.

- Support the development of long and short term requirements for database administration and design.
- Collaborate with the technical teams and the EHR vendor to ensure proactive support.
- Participate in your change control process for all planned application and technical activities.
- Work with report writers to provide data needed for reports.
- Take on full roadmap items, work with other functional teams, and be able to deliver high quality results on time.
- Investigate and find the root cause for software problems reported by clients.
- Direct organization of requirements and data into a usable database schema by directing development of ad hoc queries, scripts, macros, updates to existing queries.
- May perform database administration and maintenance, including database installation and configuration, backups, upgrades, patching.
- Document schema design in E-R diagrams using MS Visio and maintain data dictionaries for all tables.
- Review SQL code written by application developers to ensure compliance to coding standards and best practices as well as maximum performance.
- Create deployment and rollback scripts for all database objects manually or auto generated.
- Analyze access patterns and propose the best combination of indexes, constraints, foreign keys, and queries.

## **12. Role and responsibilities of a Data Analyst:**

**Answer:**Analytics brings together theory and practice to identify and communicate data-driven insights that allow managers, stakeholders, and other executives in an organization to make more informed decisions. Experienced data analysts consider their work in a larger context, within their organization and in consideration of a wide range of external factors. Analysts are also able to account for the competitive environment, internal and external business interests, and the absence of certain data sets in the data-based recommendations that they make to stakeholders.

A Master of Professional Studies in Analytics prepares students for a career as a data analyst by covering the concepts of probability theory, statistical modeling, data visualization, predictive analytics, and risk management in the context of a business environment. In addition, a master's degree in analytics equips students with the programming languages, database languages, and software programs that are vital to the day-to-day work of a data analyst.

### **Key Responsibilities of a Data Analyst**

The answer to the question “What does a data analyst do?” will vary depending on the type of organization and the extent to which a business has adopted data-driven decision-making practices. Generally speaking, though, the responsibilities data analysts must perform on a daily basis typically include the following:

- Designing and maintaining data systems and databases; this includes fixing coding errors and other data-related problems.
- Mining data from primary and secondary sources, then reorganizing said data in a format that can be easily read by either human or machine.

- Using statistical tools to interpret data sets, paying particular attention to trends and patterns that could be valuable for diagnostic and predictive analytics efforts.
- Demonstrating the significance of their work in the context of local, national, and global trends that impact both their organization and industry.
- Preparing reports for executive leadership that effectively communicate trends, patterns, and predictions using relevant data.
- Collaborating with programmers, engineers, and organizational leaders to identify opportunities for process improvements, recommend system modifications, and develop policies for data governance.
- Creating appropriate documentation that allows stakeholders to understand the steps of the data analysis process and duplicate or replicate the analysis if necessary.

### **13. Role and responsibilities of a Data Engineer:**

**Answer:** Here is the list of roles and responsibilities, Data Engineers are expected to perform:

#### **1. Work on Data Architecture**

They use a systematic approach to plan, create, and maintain data architectures while also keeping it aligned with business requirements.

#### **2. Collect Data**

Before initiating any work on the database, they have to obtain data from the right sources. After formulating a set of dataset processes, data engineers store optimized data.

#### **3. Conduct Research**

Data engineers conduct research in the industry to address any issues that can arise while tackling a business problem.

#### **4. Improve Skills**

Data engineers don't rely on theoretical database concepts alone. They must have the knowledge and prowess to work in any development environment regardless of their programming language. Similarly, they must keep themselves up-to-date with machine learning and its algorithms like the random forest, decision tree, k-means, and others.

They are proficient in analytics tools like Tableau, Knime, and Apache Spark. They use these tools to generate valuable business insights for all types of industries. For instance, data engineers can make a difference in the health industry and identify patterns in patient behavior to improve diagnosis and treatment. Similarly, law enforcement engineers can observe changes in crime rates.

#### **5. Create Models and Identify Patterns**

Data engineers use a descriptive data model for data aggregation to extract historical insights. They also make predictive models where they apply forecasting techniques to learn about the future with

actionable insights. Likewise, they utilize a prescriptive model, allowing users to take advantage of recommendations for different outcomes. A considerable chunk of a data engineer's time is spent on identifying hidden patterns from stored data.

## **6. Automate Tasks**

Data engineers dive into data and pinpoint tasks where manual participation can be eliminated with automation.

## **14. Role and responsibilities of a Data Scientist:**

**Answer:**Data scientist roles and responsibilities include:

- Data mining or extracting usable data from valuable data sources
- Using machine learning tools to select features, create and optimize classifiers
- Carrying out preprocessing of structured and unstructured data
- Enhancing data collection procedures to include all relevant information for developing analytic systems
- Processing, cleansing, and validating the integrity of data to be used for analysis
- Analyzing large amounts of information to find patterns and solutions
- Developing prediction systems and machine learning algorithms
- Presenting results in a clear manner
- Propose solutions and strategies to tackle business challenges
- Collaborate with Business and IT teams

## **15. Role and responsibilities of a Machine Learning Engineer:**

**Answer:**

- To research, modify, and apply data science and data analytics prototypes.
- To create and construct methods and plans for machine learning.
- Employing test findings to do statistical analysis and improve models.
- To search internet for training datasets that are readily available.
- ML systems and models should be trained and retrained as necessary.
- To improve and broaden current ML frameworks and libraries.
- To create machine learning applications in accordance with client or customer needs.
- To investigate, test, and put into practice appropriate ML tools and algorithms.
- To evaluate the application cases and problem-solving potential of ML algorithms and rank them according to success likelihood.
- To better comprehend data through exploration and visualization, as well as to spot discrepancies in data distribution that might affect a model's effectiveness when used in practical situations.