

### 1. What Data Science?

**Answer:** Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems to perform tasks that ordinarily require human intelligence.

### 2. What 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, analyse data, make recommendations, and more.

### 3. What is Machine Learning?

**Answer:** Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed. Machine learning focuses on developing computer programs that can access data and use it to learn for themselves.

### 4. Future of Data Science and Real-life applications

**Answer:** Let us discuss the top data science applications

- Manufacturing Process Optimization
- Medical Image Analysis
- Internet Search
- Speech Recognition
- Airline Route Planning
- Data Science in Defending Against Cyber-threats
- Data Science Application in Transportation
- Customer Behavioural Analytics
- In E-commerce
- In Healthcare
- Autocomplete
- Fraud and Risk Detection
- Targeted Advertising

**Future Of Data Science:** The future of data science is believed to witness some of the biggest innovations seen in the last decade, starting from the data explosion to the growth of the internet of things (IoT) and social media. Experts predict that in the next decade, the rise of machines will lead to the growth in usage and utility of computer systems and mobile devices.

Furthermore, experts also claim that social media use will substantially increase with users consuming numerous amounts of data online. Customers will be using social media for entertainment, transactions, surveillance, etc. Machine learning algorithms will also see a steep rise estimates some experts.

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

**Answer:** 13 Factors of Python Popularity

- 1) Easy to Learn and Use
- 2) Mature and Supportive Python Community

- 3) Support from Renowned Corporate Sponsors
  - 4) Hundreds of Python Libraries and Frameworks
  - 5) Versatility, Efficiency, Reliability, and Speed
  - 6) Big data, Machine Learning and Cloud Computing
  - 7) First-choice Language
  - 8) The Flexibility of Python Language
  - 9) Use of python in academics
  - 10) Automation
  - 11) It is interpreted
  - 12) It is open source
  - 13) Python empowers custom automation
6. Why Python for data analytics?

**Answer:** Python is a popular multi-purpose programming language widely used for its flexibility, as well as its extensive collection of libraries, which are valuable for analytics and complex calculations. Python's extensibility means that it has thousands of libraries dedicated to analytics, including the widely used Python Data Analysis Library (also known as Pandas). For the most part, data analytics libraries in Python are at least somewhat derived from the NumPy library, which includes hundreds of mathematical calculations, operations, and functions. Python analytics tools have become popular due to the computer language's widespread adoption and its versatility when it comes to developing multifaceted solutions. The fact that it's a truly general-purpose language means that it can also add deeper functionality to data analytics software than domain-specific languages that have a narrow scope and functionality. Additionally, Python's performance capability is much higher than other popular languages used in data analytics, and its compatibility with a greater array of other languages means that it is simply more convenient in most cases. Python's relatively light usage of memory and other processing resources means that it can quickly outstrip languages like MATLAB or R, which are built specifically for statistical analysis.

7. Importance of Python in machine learning.

**Answer:** Machine learning is the new black in the IT industry! All the cool kids use it – including NASA. They use machine learning to make their spacecraft travel faster and further. The rest of the gang, i.e., Uber, Tesla, and Google, also befriended it. There is at least one more thing these companies have in common. They all use python as their preferred language for their machine learning projects. Read this article to find out why.

8 reasons why Python is the preferred language for Machine Learning

- It's easy to understand and allows for quick data validation
- It has a great library ecosystem
- A low entry barrier
- It's very flexible.
- It's versatile
- It's easy to read
- Growing popularity
- Huge community

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

**Answer:** Large volumes of textual data

Natural language processing helps computers communicate with humans in their own language and scales other language-related tasks. For example, NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment, and determine which parts are important. Today's machines can analyse more language-based data than humans, without fatigue and in a consistent, unbiased way. Considering the staggering amount of unstructured data that's generated every day, from medical records to social media, automation will be critical to fully analyse text and speech data efficiently.

Structuring a highly unstructured data source

Human language is astoundingly complex and diverse. We express ourselves in infinite ways, both verbally and in writing. Not only are there hundreds of languages and dialects, but within each language is a unique set of grammar and syntax rules, terms, and slang. When we write, we often misspell or abbreviate words, or omit punctuation. When we speak, we have regional accents, and we mumble, stutter, and borrow terms from other languages.

## 9. Supervised Learning with examples.

**Answer:** Supervised Machine Learning is an algorithm that learns from labelled training data to help you predict outcomes for unforeseen data. In Supervised learning, you train the machine using data that is well "labelled." It means some data is already tagged with correct answers. It can be compared to learning in the presence of a supervisor or a teacher.

Example: Is it a cat or a dog?

Image classification is a popular problem in the computer vision field. Here, the goal is to predict what class an image belongs to. In this set of problems, we are interested in finding the class label of an image. More precisely: is the image of a car or a plane? A cat or a dog?

## 10. Unsupervised learning with examples.

**Answer:** Unsupervised learning is when it can provide a set of unlabelled data, which it is required to analyse and find patterns inside. The examples are dimension reduction and clustering.

Here is an example of a real-world problem solved using unsupervised learning on satellite images to identify climate anomalies. Somalia is a small country in the continent of Africa. The country exhibits a lot of natural disasters and terrorism because of which people of Somalia go through mass displacements leading towards a situation of lack of food and shelter. This article shows how to build an anomaly detection system using Machine Learning. The system can capture sudden vegetation changes, which can be used as an alert mechanism to provide immediate relief to the people and communities in need.

## Job:

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

**Answer:** Here are examples of responsibilities from real database engineer resumes representing typical tasks they are likely to perform in their roles.

- Create PHP page for loan officers to add to their contact list for leads which interact with the MySQL database.
- Develop Perl and awk scripts to validate the various data on LINUX base system.
- Develop procedures; UNIX shell, SQL, and Perl scripts; and software programs for database processing.
- Trace Auto Trader Latino inventory process through the ETL and refresh to identify, troubleshoot, and correct code problems.
- Develop, implement, and execute a quality assurance program and quality control standards for all ETL (SSIS) activities.
- Pioneer mainframe development the on PC.
- Perform work in the systems architecture area.
- Assist programmers in debugging issues and finding solutions.
- Create RAC database using ASM raw disks and on NetApp filer.
- Utilize macros and VBA queries to efficiently manipulate table and query data.
- Clean up and rewrite front end HTML and CSS to meet validation requirements.
- Develop, execute, and complete a key data migration project (Perl).
- Work on deletion prediction project using shell scripting, Perl scripting and Berkeley databases.
- Run explain plans to troubleshoot queries or determine the best strategy to enhance their execution.
- Adhere to document change control by raising RFC and perform modifications to databases during weekend windows.

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

**Answer:** A data analyst is responsible for organizing data related to sales numbers, market research, logistics, linguistics, or other behaviours. They utilize technical expertise to ensure data is accurate and high-quality. Data is then analysed, designed, and presented in a way that assists individuals, businesses, and organizations make better decisions.

- Using automated tools to extract data from primary and secondary sources
- Removing corrupted data and fixing coding errors and related problems
- Developing and maintaining databases, data systems – reorganizing data in a readable format
- Performing analysis to assess quality and meaning of data
- Filter Data by reviewing reports and performance indicators to identify and correct code problems
- Using statistical tools to identify, analyse, and interpret patterns and trends in complex data sets that could be helpful for the diagnosis and prediction
- Assigning numerical value to essential business functions so that business performance can be assessed and compared over periods of time.
- Analysing local, national, and global trends that impact both the organization and the industry

- Preparing reports for the management stating trends, patterns, and predictions using relevant data
- Working with programmers, engineers, and management heads to identify process improvement opportunities, propose system modifications, and devise data governance strategies.
- Preparing final analysis reports for the stakeholders to understand the data-analysis steps, enabling them to take important decisions based on various facts and trends.

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

**Answer:** Regardless of the focus on a specific part of a system, data engineers have similar responsibilities. This is mostly a technical position that combines knowledge and skills of computer science, engineering, and databases.

- Architecture design.** In its core, data engineering entails designing the architecture of a data platform.
- Development of data related instruments/instances.** As a data engineer is a developer role in the first place, these specialists use programming skills to develop, customize and manage integration tools, databases, warehouses, and analytical systems.
- Data pipeline maintenance/testing.** During the development phase, data engineers would test the reliability and performance of each part of a system. Or they can cooperate with the testing team.
- Machine learning algorithm deployment.** Machine learning models are designed by data scientists. Data engineers are responsible for deploying those into production environments. This entails providing the model with data stored in a warehouse or coming directly from sources, configuring data attributes, managing computing resources, setting up monitoring tools, etc.
- Manage data and meta-data.** The data can be stored in a warehouse either in a structured or unstructured way. Additional storage may contain meta-data (exploratory data about data). A data engineer is in charge of managing the data stored and structuring it properly via database management systems.
- Provide data-access tools.** In some cases, such tools are not required, as warehouse types like data-lakes can be used by data scientists to pull data right from storage. However, if an organization requires business intelligence for analysts and other non-technical users, data engineers are responsible for setting up tools to view data, generate reports, and create visuals.
- Track pipeline stability.** Monitoring the overall performance and stability of the system is important if the warehouse needs to be cleaned from time to time. The automated parts of a pipeline should also be monitored and modified since data/models/requirements can change.

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

**Answer:** Data scientists work closely with business stakeholders to understand their goals and determine how data can be used to achieve those goals. They design data modelling processes, create algorithms and predictive models to extract the data the business needs, and help analyse the data and share insights with peers. While each project is different, the process for gathering and analysing data generally follows the below path:

1. Ask the right questions to begin the discovery process

2. Acquire data
3. Process and clean the data
4. Integrate and store data
5. Initial data investigation and exploratory data analysis
6. Choose one or more potential models and algorithms
7. Apply data science techniques, such as machine learning, statistical modelling, and artificial intelligence
8. Measure and improve results
9. Present final result to stakeholders
10. Make adjustments based on feedback
11. Repeat the process to solve a new problem

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.

**Skills of an ML Engineer**

A person who wants to work as a machine learning engineer needs to possess the following skills and credentials:

- Advanced math and statistics knowledge, particularly in the areas of calculus, linear algebra, and Bayesian statistics.
- Advanced degree in math, computer science, statistics or a related field.
- A master's degree in artificial intelligence, deep learning, or a related discipline.
- Strong teamwork, problem-solving, and analytical skills.
- Abilities in software engineering.
- Knowledge of data science.
- Languages for coding and programming, such as Python, Java, C++, C, R, and JavaScript.
- Practical understanding of ML frameworks.
- Practical familiarity with ML libraries and packages.
- Recognize software architecture, data modelling, and data structures.
- Understanding of computer architecture.