

INTRODUCTION TO DATA SCIENCE

1A

INTRODUCTION TO DATA SCIENCE



INTRODUCTION TO DATA SCIENCE

INTRODUCTION

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 is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data. Data science is related to data mining, machine learning and big data.

INTRODUCTION TO DATA SCIENCE

Steps to Become a Data Scientist / Analysts

- Brush up on your math and statistics skills
- Learn to code
- Understand databases, data lakes and distributed storage
- Learn data munging and data cleaning techniques
- Understand the basics of good data visualisation and reporting
- Understand the concept of machine learning
- Add more tools to your toolbox (Python, R, Tableau, QlikView. SPSS etc.)
- Lots of Practice

INTRODUCTION TO DATA SCIENCE

What is the most important thing in data science?

The most important things to learn in **Data Science** are: Mathematical concepts such as **linear algebra, probabilities, and distributions**. Statistical concepts such as **descriptive** and **inferential statistics**. Programming languages such as **Python, R, and SAS**.

INTRODUCTION TO DATA SCIENCE

What are the different types of data scientists?

Different types of Data Scientists

- 1) Data Scientist as Statistician
- 2) Data Scientist as Mathematician
- 3) Data Scientists Vs Data Engineers
- 4) Data Scientist as Machine Learning Scientists
- 5) Data Scientist as Actuarial Scientist
- 6) Data Scientist as Business Analytic Practitioners
- 7) Data Scientist as Software Programming Analysts

INTRODUCTION TO DATA SCIENCE

What is the focus of data science?

Data Science is a more **forward-looking** approach, an exploratory way with the focus on analyzing the **past** or current data and **predicting** the **future** outcomes with the aim of making **informed decisions**. It answers the open-ended questions as to "**what**" and "**how**" events occur.

INTRODUCTION TO DATA SCIENCE

What is the first step in data science process?

The Data Science Process

Step 1: Frame the problem. The first thing you have to do before you solve a problem is to define exactly what it is....

Step 2: Collect the raw data needed for your problem....

Step 3: Process the data for analysis....

Step 4: Explore the data....

Step 5: Perform in-depth analysis....

Step 6: Communicate results of the analysis.

INTRODUCTION TO DATA SCIENCE

What are the basics of data science?

The Introduction to Data Science class will survey the foundational topics in data science, namely:

- Data **Manipulation**.
- Data **Analysis** with **Statistics** and **Machine Learning**.
- Data **Communication** with **Information Visualization**.

INTRODUCTION TO DATA SCIENCE

What is the principal goal of data science?

The **goal** of data science is to construct the means for extracting business-focused insights from **data**. This requires an understanding of how **value** and **information** flows in a business, and the ability to use that understanding to identify **business opportunities**.

INTRODUCTION TO DATA SCIENCE

Does data science require coding?

You need to have the knowledge of programming languages like **Python**, **Perl**, **R**, **SQL**, and **Java**—with **Python** being the most common coding language required in data science roles. Programming languages help you **clean**, **massage**, and **organize** an **unstructured** set of data.

Why is Python necessary for data science?

Python is the most common coding language required in data science roles, along with Java, or R programming. Python is a great programming language for **data scientists**. Because of its **versatility**, we can use **Python** for almost all the steps involved in data science processes.

INTRODUCTION TO DATA SCIENCE

What is the benefit of data science?

Data Science training empowers professionals with data management technologies such as **Hadoop**, **Flume**, **Machine learning**, etc. If a candidate has the knowledge and proficiency of these significant data skills, it would be an added advantage for them to have an **improved** and **competitive** career.

INTRODUCTION TO DATA SCIENCE

5 Top Industries for Data Science Professionals

- Healthcare & Pharmaceutical. Building a career in the healthcare sector is a delicate matter
- Telecommunications Sector
- Internet Industry
- Energy Sector
- Automotive Industry

INTRODUCTION TO DATA SCIENCE

Which industry or field will most benefit from data science?

Industries That Benefits the Most From Data Science

- Construction
- Retail
- Medicine
- Transportation
- Communications, Media, and Entertainment
- Education
- Manufacturing and Natural Resources
- Government
- Energy and Utilities
- Outsourcing Industry

INTRODUCTION TO DATA SCIENCE

Here is the list of best data science tools that is mostly used in the industries.

- SAS
- Apache Spark
- BigML
- MATLAB
- Excel
- ggplot2
- Tableau
- Jupyter
- Matplotlib
- NLTL
- Scikit-learn
- TensorFlow
- Weka

INTRODUCTION TO DATA SCIENCE

THANK YOU