

IBM DS0101EN

What is Data Science?

Instructor Name

- MACHINE LEARNING INTERN, IBM EGYPT
- MACHINE LEARNING ENGINEER at ABC
- MACHINE LEARNING ENGINEER at ABD
- PYTHON AND DATA SCIENCE INSTRUCTOR

- In this course, you'll learn how to:
 - Define data science and its importance in today's data-driven world.
 - Describe the various paths that can lead to a career in data science.
 - Summarize advice given by seasoned data science professionals to data scientists who are just starting out.
- Course Modules:
 - **Define data science & What Data Scientists Do**
 - **Data science Topics**
 - **Applications and Careers in Data Science**
 - **Data Literacy for Data Science**

Module I

Define data science & What Data Scientists Do



Module I

Define data science & What Data Scientists Do

What is Data Science?

Understanding Data Science

- Data Science is a continuous process of utilizing data to gain insights.
- It involves validating hypotheses or models using available data.
- The goal is to uncover trends and insights hidden within datasets.

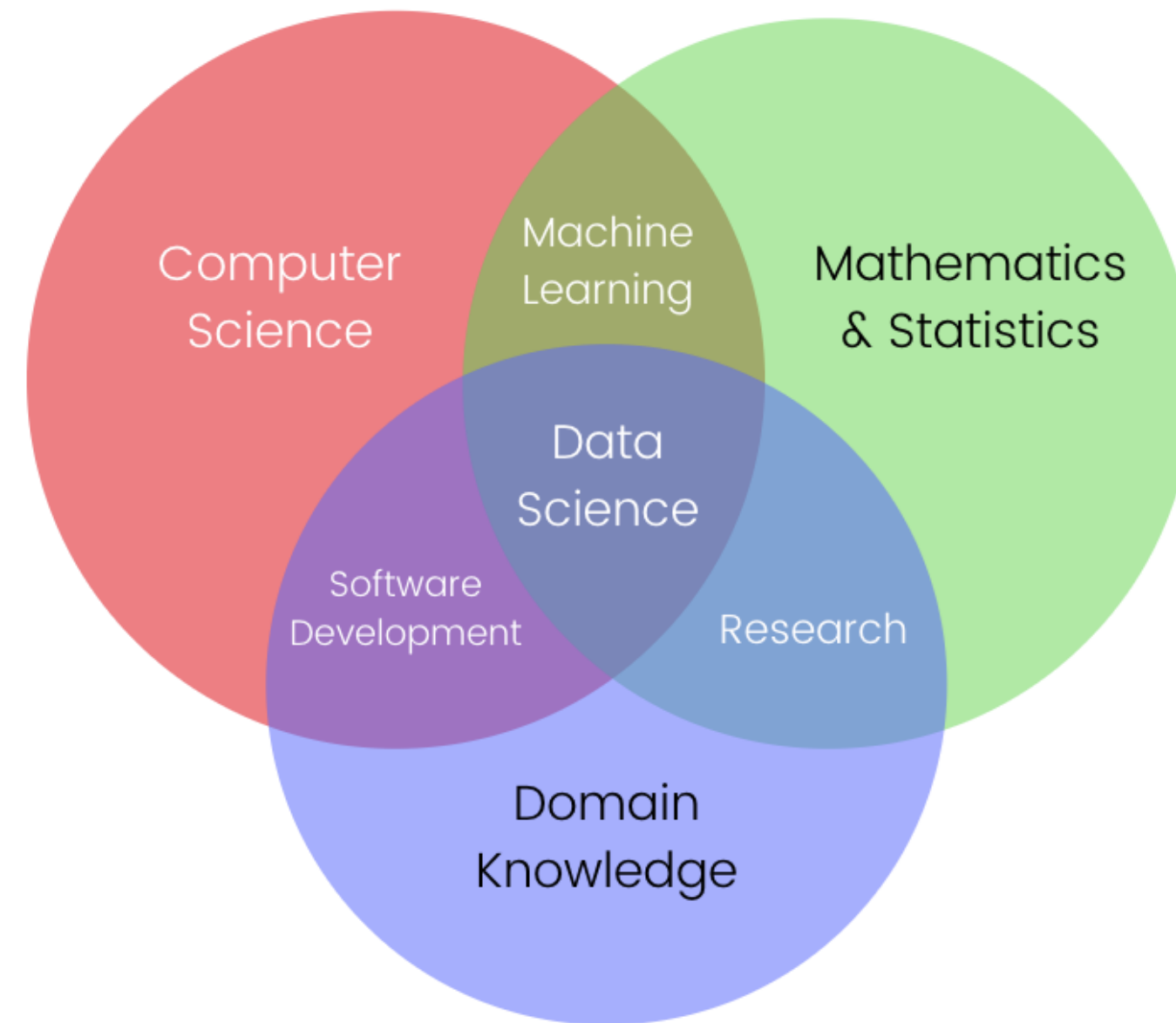


Understanding Data Science

- Data is transformed into compelling narratives through storytelling.
- These insights drive strategic decision-making for organizations.
- It encompasses extracting and analyzing data in structured and unstructured forms.



Understanding Data Science



The Essence of Data Science

- Data science explores, manipulates, and analyzes data to find answers.
- Just as other sciences study specific subjects, data science focuses on understanding data.
- Today's world offers an abundance of data, algorithms, and accessible tools.
- The affordability and accessibility of these resources make data science more relevant than ever.



The Essence of Data Science

- It's a time of unprecedented opportunity for those interested in data science.
- Data science thrives on curiosity, exploration, and leveraging available resources.



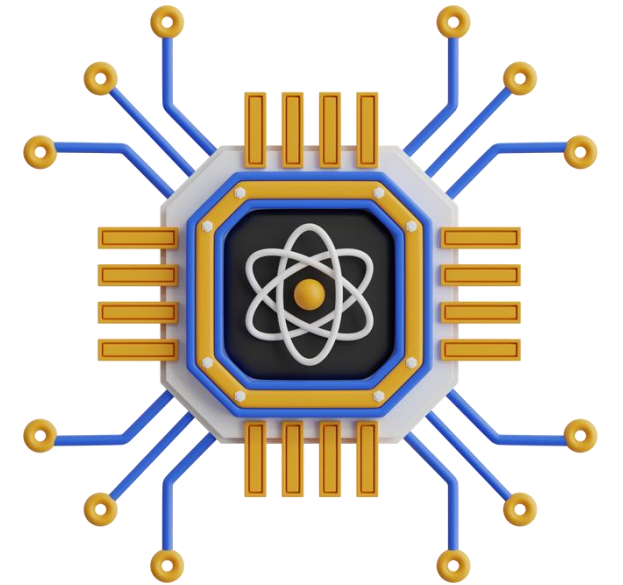
Module I

Define data science & What Data Scientists Do

Fundamentals of Data Science

Understanding Data Science

- Data Science encompasses significant data analysis across various sources.
- It leverages vast quantities of data from diverse sources like social media and sales.
- Advancements in computing power enable meaningful analysis and new discoveries.

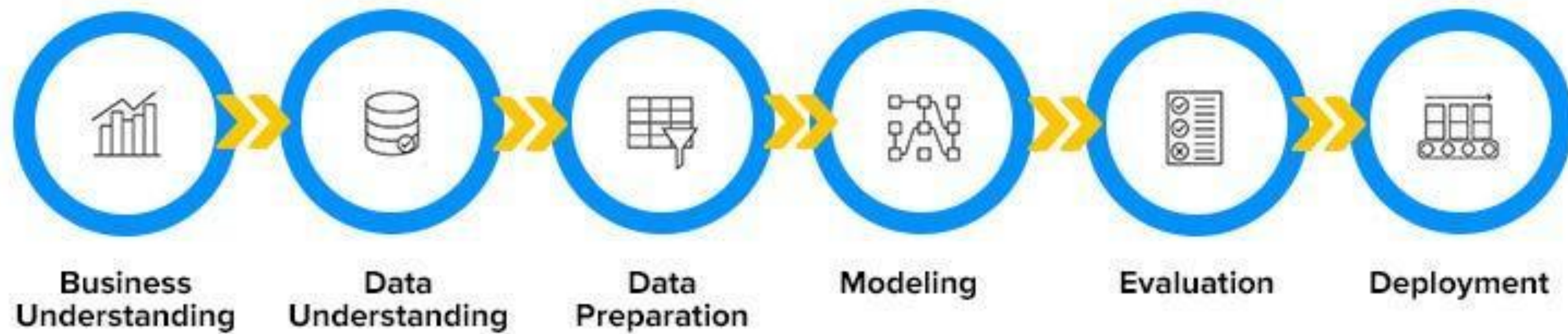


Understanding Data Science

- Data science aids organizations in understanding their environments and uncovering opportunities.
- Data scientists investigate data to add value and insight to the organization's knowledge.
- The process starts with clarifying the organization's question or problem.



The Data Science Process



Module I

Define data science & What Data Scientists Do

Real-Life Applications of Data Science

REAL-LIFE APPLICATIONS OF DATA SCIENCE

Let's see some use cases



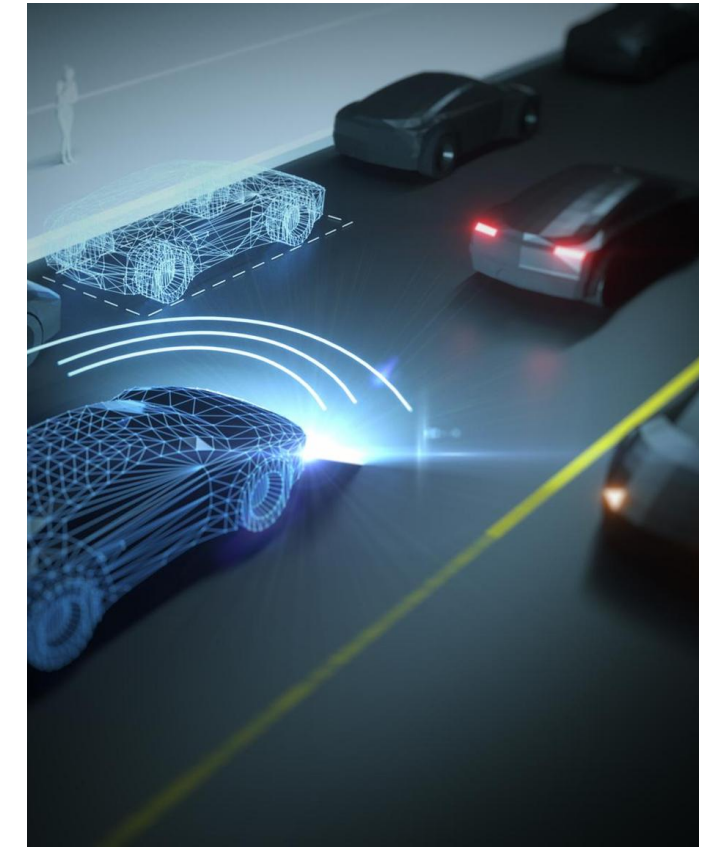
CHATBOT



VOICE ASSISTANTS



RECOMMENDATION SYSTEM



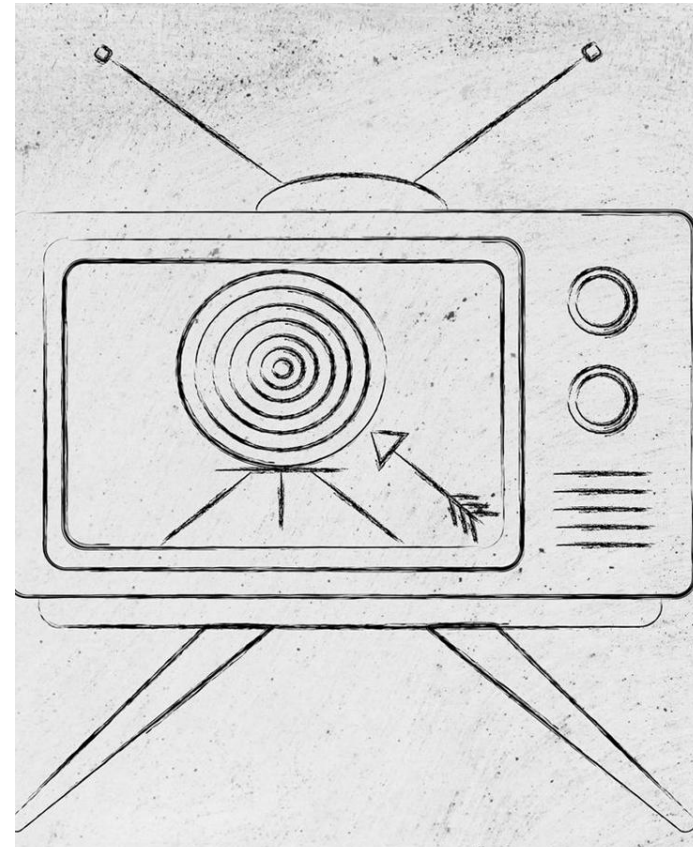
SELF-DRIVING CARS

REAL-LIFE APPLICATIONS OF DATA SCIENCE

Let's see some use cases



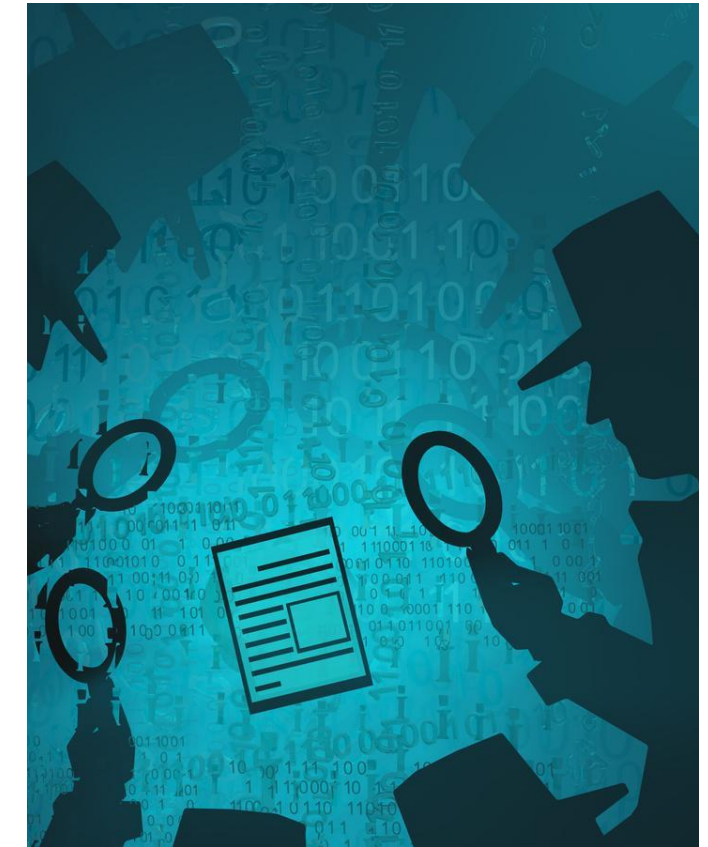
CHESS



ADS TARGETING



HEALTH CARE



FRAUD DETECTION

Module I

Define data science & What Data Scientists Do

The Many Paths to Data Science

Evolution of Data Science Careers

- Data science was not a recognized field until around 2009-2011.
- DJ Patil and Andrew Gelman are credited with coining the term.
- Before data science, statistics was a prevalent field.
- Individuals often pursued business or other quantitative analysis disciplines.
- Exposure to data science often occurred during academic or professional endeavors.
- The term "data science" gained prominence in various industries over time.

Personal Journeys into Data Science

- Many individuals stumbled into data science through academic or professional paths.
- Backgrounds varied from engineering to business, economics, and analytics.
- Exposure to data science often occurred during higher education or internships.
- Practical applications in fields like transportation engineering introduced individuals to data science.
- Gradual immersion in data analysis and modeling paved the way for careers in data science.
- The journey into data science showcases diverse paths and backgrounds converging into the field.

Module I

Define data science & What Data Scientists Do

Advice for New Data Scientists

Advice FOR NEW DATA SCIENTISTS

01

Curiosity

Curiosity is fundamental for exploring and understanding complex data.

02

Being judgmental

Being judgmental helps in forming hypotheses and initial assumptions.

03

Argumentativeness

Argumentativeness aids in advocating for a specific direction and learning from data.

04

Flexibility

Comfort and flexibility with analytics platforms are valuable secondary skills.

Career Development Strategies for Data Scientists

- Identify your competitive advantage and preferred industry focus.
- Tailor your analytical skills to match the needs of your chosen field.
- Acquire proficiency in industry-specific analytics platforms and tools.
- Apply your skills to real-world problems to demonstrate your capabilities.
- Develop storytelling abilities to effectively communicate insights and findings.
- Continuously refine and adapt your skills to stay relevant and competitive in the field.

Module I

Define data science & What Data Scientists Do

Lesson Summary: Defining Data Science

Understanding Data Science

- Data science studies data to understand the world around us.
- It uncovers insights and trends hidden within vast amounts of data.
- Recent advancements in computing power enable deeper analysis and new knowledge.
- Data scientists play a crucial role in translating data into actionable insights.
- The process involves problem clarification, data collection, analysis, and visualization.
- Curiosity, argumentation, and judgment are key traits for successful data scientists.

Developing Skills and Career Paths

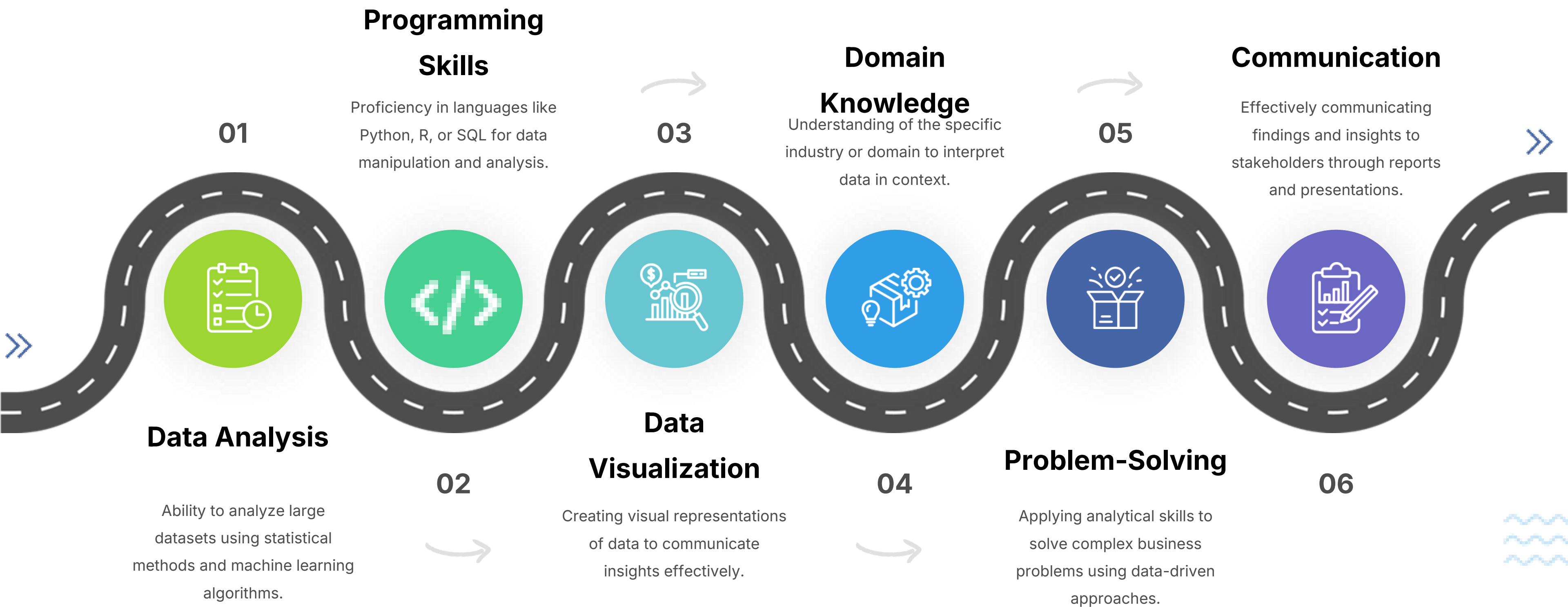
- Skilled data scientists possess versatile knowledge beyond statistics and programming.
- They come from diverse backgrounds such as economics, engineering, or medicine.
- Mastery of data analysis tools and techniques is essential for success.
- Specialization in a particular field enhances expertise and industry relevance.
- Certification may become necessary as companies prioritize qualified candidates.
- Future data scientists will adapt to evolving technology and changing job roles for successful business outcomes.

Module I

Define data science & What Data Scientists Do

Data Science & Big Data Skills

Data Scientist Skills

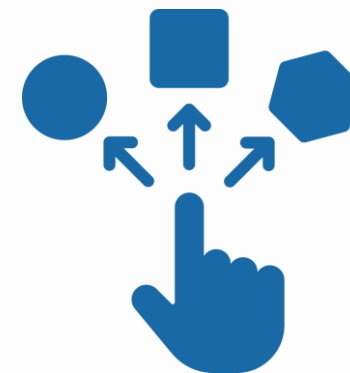


Big Data



Volume

Dealing with large volumes of data that traditional systems cannot handle efficiently.



Variety

Handling diverse data types such as structured, semi-structured, and unstructured data.



Velocity

Processing data streams in real-time or near real-time to derive timely insights.

Module I

Define data science & What Data Scientists Do

Understanding Different Types of File Formats

Understanding File Formats

- Data professionals work with various file types and formats.
- Importance of understanding file structure, benefits, and limitations.
- Choosing suitable formats for data and performance requirements.
- Covered file formats: Delimited text, XLSX, XML, PDF, JSON.
- Delimited text files: Rows with values separated by delimiters like commas or tabs.
- CSVs and TSVs are common in this category and suited for straightforward information.

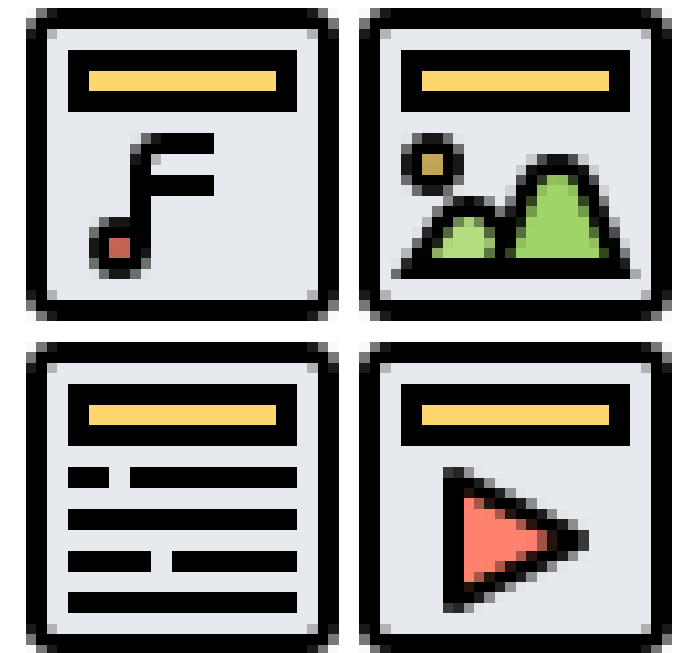
Overview of File Formats

XLSX:

- Microsoft Excel's Open XML format.
- Data is organized into worksheets (rows and columns).
- Supports formulas, charts, and data manipulation.
- Widely used for data analysis and reporting.

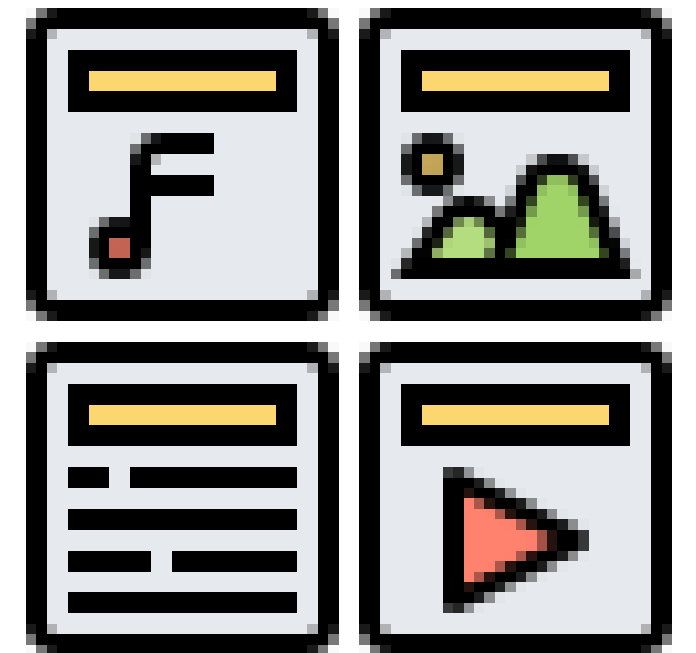
XML:

- A markup language designed to encode data.
- Both human- and machine-readable.
- Often used for data exchange between systems and applications.
- Flexible and customizable structure for storing and transmitting data.



Overview of File Formats

- **PDF:**
- Developed by Adobe for consistent document presentation.
- Maintains formatting across different devices and platforms.
- Ideal for sharing documents where layout, fonts, and images must remain intact.
- Commonly used for reports, contracts, and presentations.



Overview of File Formats

- JSON: Text-based standard for transmitting structured data over the web.
- JSON is language-independent, easy to use, and widely compatible.

Understanding popular file formats is crucial for effective data handling and analysis.

Example of File Formats

JSON

```
{  
  "siblings": [  
    {"firstName": "Anna", "lastName": "Clayton"},  
    {"firstName": "Alex", "lastName": "Clayton"}  
  ]  
}
```

XML

```
<siblings>  
  <sibling>  
    <firstName>Anna</firstName>  
    <lastName>Clayton</lastName>  
  </sibling>  
  <sibling>  
    <firstName>Alex</firstName>  
    <lastName>Clayton</lastName>  
  </sibling>  
</siblings>
```

Q & A

Thank you!