IBM Data Science Certificate – Coursera – Started 12/02/2025

**Defining Data Science and What Data Scientists Do**

**Defining Data Science**

* Defining Data Science
* Video: What is Data Science?
* Fundamentals of Data Science
* The Many Paths to Data Science
* Data Science: The Sexiest Job in the 21st Century
* Defining Data Science
* Advice for New Data Scientists

**What Do Data Scientists Do?**

* A Day in the Life of a Data Scientist
* Data Science Skills & Big Data
* Working on Different File Formats
* Data Science Topics and Algorithms
* Discussion Prompt: Introduce Yourself
* Reading: What Makes Someone a Data Scientist?

**Data Science Topics**

**Big Data and Data Mining**

* How Big Data is Driving Digital Transformation
* Introduction to Cloud
* Cloud for Data Science
* Foundations of Big Data
* Data Scientists at New York University
* What is Hadoop?
* Big Data Processing Tools: Hadoop, HDFS, Hive, and Spark
* Reading: Data Mining

**Deep Learning and Machine Learning**

* Artificial Intelligence and Data Science
* Generative AI and Data Science
* Neural Networks and Deep Learning
* Applications of Machine Learning
* Reading: Regression
* Lab: Exploring Data using IBM Cloud Gallery

**Applications and Careers in Data Science:**

**Data Science Application Domains**

* How Should Companies Get Started in Data Science?
* Old Problems with New Data Science Solutions
* Applications of Data Science
* How Data Science is Saving Lives
* Reading: The Final Deliverable

**Careers and Recruiting in Data Science**

* How Can Someone Become a Data Scientist?
* Recruiting for Data Science
* Careers in Data Science
* Importance of Mathematics and Statistics for Data Science (only name change)
* The Report Structure
* Reading: Infograph on roadmap

**Data Literacy for Data Science (Optional):**

**Understanding Data**

* Understanding Data
* Data Sources
* Working on Varied Data Sources and Types
* Reading: Metadata

**Data Literacy**

* Data Collection and Organization
* Relational Database Management System
* NoSQL
* Data Marts, Data Lakes, ETL, and Data Pipelines
* Considerations for Choice of Data Repository
* Data Integration Platforms

**What is Data Science?**

Data Science is a process, not an event.

It is the process of using data to understand different things,

to understand the world.

For me is when you have a model or hypothesis of a problem,

and you try to validate that hypothesis or model with your data.

Data science is the art of

uncovering the insights and trends that are hiding behind data.

It's when you translate data into a story.

So use storytelling to generate insight.

And with these insights,

you can make strategic choices for a company or an institution.

Data science is a field about processes and systems to extract

data from various forms of whether it is unstructured or structured form.

Data science is the study of data.

Like biological sciences is a study of biology,

physical sciences, it's the study of physical reactions.

Data is real, data has real properties,

and we need to study them if we're going to work on them.

Data Science involves data and some science.

The definition or the name came up in

the 80s and 90s when some professors were looking into the statistics curriculum,

and they thought it would be better to call it data science.

But what is Data Science?

I'd see data science as one's attempt to work with data,

to find answers to questions that they are exploring.

In a nutshell, it's more about data than it is about science.

If you have data, and you have curiosity,

and you're working with data,

and you're manipulating it, you're exploring it,

the very exercise of going through analyzing data,

trying to get some answers from it is data science.

Data science is relevant today because we have tons of data available.

We used to worry about lack of data.

Now we have a data deluge.

In the past, we didn't have algorithms, now we have algorithms.

In the past, the software was expensive,

now it's open source and free.

In the past, we couldn't store large amounts of data,

now for a fraction of the cost,

we can have gazillions of datasets for a very low cost.

So, the tools to work with data,

the very availability of data,

and the ability to store and analyze data,

it's all cheap, it's all available,

it's all ubiquitous, it's here.

There's never been a better time to be a data scientist.

**Fundamentals of Data Science**

Everyone you ask will give you a slightly different description of what Data Science is, but most people agree that it has a significant data analysis component. Data analysis isn't new. What is new is the vast quantity of data available from massively varied sources: from log files, email, social media, sales data, patient information files, sports performance data, sensor data, security cameras, and many more besides. At the same time that there is more data available than ever, we have the computing power needed to make a useful analysis and reveal new knowledge. Data science can help organizations understand their environments, analyze existing issues, and reveal previously hidden opportunities. Data scientists use data analysis to add to the knowledge of the organization by investigating data, exploring the best way to use it to provide value to the business. So, what is the process of data science? Many organizations will use data science to focus on a specific problem, and so it's essential to clarify the question that the organization wants answered. This first and most crucial step defines how the data science project progresses. Good data scientists are curious people who ask questions to clarify the business need. The next questions are: "what data do we need to solve the problem, and where will that data come from?". Data scientists can analyze structured and unstructured data from many sources, and depending on the nature of the problem, they can choose to analyze the data in different ways. Using multiple models to explore the data reveals patterns and outliers; sometimes, this will confirm what the organization suspects, but sometimes it will be completely new knowledge, leading the organization to a new approach. When the data has revealed its insights, the role of the data scientist becomes that of a storyteller, communicating the results to the project stakeholders. Data scientists can use powerful data visualization tools to help stakeholders understand the nature of the results, and the recommended action to take. Data Science is changing the way we work; it's changing the way we use data and it’s changing the way organisations understand the world.

**Advice for New Data Scientists**

My advice to an aspiring data scientist is to be curious, extremely argumentative and judgmental. Curiosity is absolute must. If you're not curious, you would not know what to do with the data. Judgmental because if you do not have preconceived notions about things you wouldn't know where to begin with. Argumentative because if you can argument and if you can plead a case, at least you can start somewhere and then you learn from data and then you modify your assumptions and hypotheses and your data would help you learn. And you may start at the wrong point. You may say that I thought I believed this, but now with data I know this. So, this allows you a learning process. So, curiosity being able to take a position, strong position, and then moving forward with it. The other thing that the data scientist [should] would need is some comfort and flexibility with analytics platforms: some software, some computing platform, but that's secondary. The most important thing is curiosity and the ability to take positions. Once you have done that, once you've analyzed, then you've got some answers. And that's the last thing that a data scientist need, and that is the ability to tell a story. That once you have your analytics, once you have your tabulations, now you should be able to tell a great story from it. Because if you don't tell a great story from it, your findings will remain hidden, remain buried, nobody would know. Your rise to prominence is pretty much relying on your ability to tell great stories. A starting point would be to see what is your competitive advantage. Do you want to be a data scientist in any field or a specific field? Because, let's say you want to be a data scientist and work for an IT firm or a web-based or Internet based firm, then you need a different set of skills. And if you want to be a data scientist, for lets say, in the health industry, then you need different sets of skills. So figure out first what you're interested, and what is your competitive advantage. Your competitive advantage is not necessarily going to be your analytical skills. Your competitive advantage is your understanding of some aspect of life where you exceed beyond others in understanding that. Maybe it's film, maybe it's retail, maybe it's health, maybe it's computers. Once you've figured out where your expertise lies, then you start acquiring analytical skills. What platforms to learn and those platforms, those tools would be specific to the industry that you're interested in. And then once you have got some proficiency in the tools, the next thing would be to apply your skills to real problems, and then tell the rest of the world what you can do with it. [Music]

Key Takeaways

Welcome to defining data science, lesson summary. In this video, we'll review important points you learned from the videos and readings in this lesson, we'll also link together the ideas from the resources in this lesson. We will quickly recap what data science is, data scientists role in an organization, what makes a skilled data a scientist, and expert advice on how to acquire these skills? Let's begin, so what really is data science? Simply put, data science is the study of data, data science uses data to understand the world around us, some consider data science an art of uncovering the insights and trends hiding behind data. Data analysis isn't new, recent data access and enhanced computing power drives new insights and knowledge through analysis. We also have the computing power needed to analyze this data to reveal new knowledge, with the power of computers, we can dig through this information to reveal valuable insights. Just like a detective uncovering secrets, data scientists translate the data into stories to generate insights, these insights aid strategic decision-making for companies or institutions. Similar to biological or physical sciences, data science deals with structured and unstructured data. The process of gleaning insights from data includes clarifying the problem, data collection, analysis, pattern recognition, storytelling, and visualization. According to Professor Murtez Aheter from the Ted Rogers School of Management, curiosity, argumentation, and judgment are vital for data scientists. Curiosity helps you explore data and come up with meaningful questions. Good, sound, reasonable arguments help you explain your findings after learning from the data, compelling the listener to adjust their ideas based on the new information. Good judgment guides you to start in the right direction. Skilled data scientists go beyond just being statisticians or computer experts. Companies are looking for versatile individuals who know a lot about a particular subject, have some experience in programming and analyzing data, and can communicate well. Generally, data scientists are comfortable with math, they're curious, they're good at telling stories. Their backgrounds can come from various fields like economics, engineering, medicine, and more. Once you understand your strengths and interests, focus on mastering data analysis in that field and select suitable tools for your industry. As you become proficient, apply your expertise to solve real world issues using data, similar to solving mysteries by deciphering clues. So, what does the future look like for you as a skilled data scientist? Data scientist jobs will also change as technology changes and data roles develop, to ensure their employees are qualified, companies will require certification. Data scientists will always need to think logically, use algorithms, and follow a methodical approach. Most importantly, they must gather data correctly and carefully analyze the models being used, all aiming to achieve successful business results. [MUSIC]

**Defining Data Science Lesson Glossary**

Welcome! This alphabetized glossary contains many of the terms in this course. These terms are important for you to recognize when working in the industry, participating in user groups, and participating in other certificate programs.

* Term Definition Video where the term is introduced
* Algorithms A set of step-by-step instructions to solve a problem or complete a task. What is Data Science?
* Model A representation of the relationships and patterns found in data to make predictions or analyze complex systems retaining essential elements needed for analysis. What is Data Science?
* Outliers When a data point or points occur significantly outside of most of the other data in a data set, potentially indicating anomalies, errors, or unique phenomena that could impact statistical analysis or modeling. What is Data Science?
* Quantitative analysis A systematic approach using mathematical and statistical analysis is used to interpret numerical data. Many Paths to Data Science
* Structured data Data is organized and formatted into a predictable schema, usually related tables with rows and columns. What is Data Science?
* Unstructured data Unorganized data that lacks a predefined data model or organization makes it harder to analyze using traditional methods. This data type often includes text, images, videos, and other content that doesn’t fit neatly into rows and columns like structured data. What is Data Science?hare