



LOVELY
PROFESSIONAL
UNIVERSITY

Report on

WHAT IS DATA SCIENCE?

Submitted to

LOVELY PROFESSIONAL UNIVERSITY

in partial fulfilment of the requirements for the award of degree of

Master of Computer Application

FROM 25 MAY 2022 TO 28 JUNE 2022

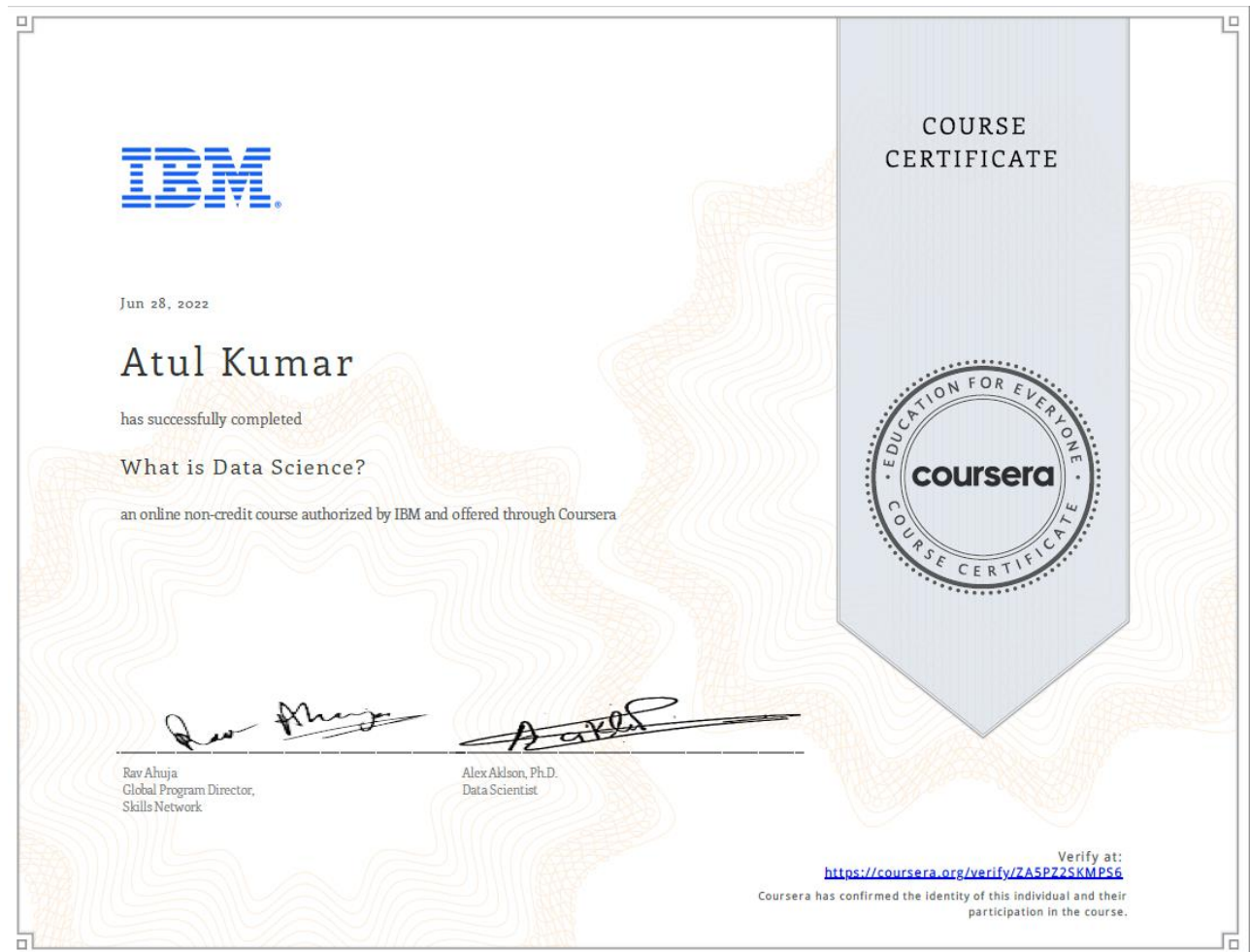
Submitted By

Atul Kumar

Supervised By

**Mr. Kumar Vishal
Professor**

CERTIFICATE



Name: What is Data Science?

Issuing Organization: Coursera

Issue Date: June 2022

Expiration Date: *This certification does not expire*

Credential ID: ZA5PZ2SKMPS6

Credential URL: <https://www.coursera.org/account/accomplishments/certificate/ZA5PZ2SKMPS6>

DECLARATION

We hereby declare that the **MOOC** Course work entitled **What is Data Science** is an authentic record of our own work carried out as requirements of Course for the award of master's degree of Computer Application and Engineering from Lovely Professional University, Phagwara, under the guidance of Mr. Kumar Vishal, during MAY to JUNE 2022. All the information furnished in this **MOOC** Course report is based on our own intensive work and is genuine.

Atul Kumar [12102801]

CERTIFICATE

This is to certify that ATUL KUMAR [12102801] have completed their Course Certificate titled **What is Data Science** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original study. No part of the report has ever been submitted for any other degree or diploma. The report is fit for the submission and the partial fulfilment of the conditions for the award of MCA.

23 September 2022

Mr. Kumar Vishal

TABLE OF CONTENTS

INTRODUCTION

WEEK 1

1. Course Syllabus
 - 1.1 Professional Certificate Career Support
2. Defining Data Science
 - 2.1 What is data Science
 - 2.2 Fundamentals of Data Science
 - 2.3 The Many Paths to Data Science
 - 2.4 Advice for New Data Scientists
 - 2.5 The Sexiest Job in the 21st Century
 - 2.6 The Sexiest Job in the 21st Century (Quiz)
 - 2.7 Lesson Summary
3. What Do Data Scientists Do?
 - 3.1 A day in the Life of a Data Scientist
 - 3.2 Old problems, new problems, Data Science Solutions
 - 3.3 Data Science Topics and Algorithms
 - 3.4 Introduce Yourself
 - 3.5 What Makes Someone a Data Scientist?
 - 3.6 What Makes Someone a Data Scientist? (Quiz)
 - 3.7 Lesson Summary
 - 3.8 Learning Outcomes

WEEK 2

1. Big Data and Data Mining
 - 1.1 Foundation of Big Data
 - 1.2 What is Hadoop?
 - 1.3 How Big Data is Driving Digital Transformation
 - 1.4 Data Science Skills & Big Data
 - 1.5 Data Mining

- 1.6 Data Mining (Graded Quiz)
- 1.7 Lesson Summary
- 2. Deep Learning and Machine Learning
 - 2.1 Neural Networks and Deep Learning
 - 2.2 Applications of Machine Learning
 - 2.3 Regression
 - 2.4 Regression (Graded Quiz)
 - 2.5 Lesson Summary
- 3. Hands-on Exercise: Data Science Exploration
 - 3.1 Obtain an IBM Cloud Feature Code
 - 3.2 Lab: Create IBM Cloud Account
 - 3.3 Lab: Create Watson Studio Instance
 - 3.4 Learning Outcomes

WEEK 3

- 1. Data Science in Business
 - 1.1 How Data Science is Saving Lives
 - 1.2 How Should Companies Get Started in Data Science?
 - 1.3 Applications of Data Science
 - 1.4 The Final Deliverable
 - 1.5 The Final Deliverable (Quiz)
 - 1.6 Lesson Summary
- 2. Careers and Recruiting in Data Science
 - 2.1 How can Someone Becomes a data Scientist
 - 2.2 Recruiting for Data Science
 - 2.3 Careers in Data Science
 - 2.4 High School Students and Data Science Careers
 - 2.5 Lesson Summary
- 3. Data Science in Business
 - 3.1 How Data Science is Saving Lives
 - 3.2 How Should Companies Get Started in Data Science?
 - 3.3 Applications of Data Science
 - 3.4 The Final Deliverable

3.5 The Final Deliverable (Quiz)

3.6 Lesson Summary

4. Data Science in Business

4.1 How Data Science is Saving Lives

4.2 How Should Companies Get Started in Data Science?

4.3 Applications of Data Science

4.4 The Final Deliverable

4.5 The Final Deliverable (Quiz)

4.6 Lesson Summary

5. Careers and Recruiting in Data Science

5.1 How can Someone Become a data Scientist

5.2 Recruiting for Data Science

5.3 Careers in Data Science

5.4 High School Students and Data Science Careers

5.5 Lesson Summary

Week 4

1. The Report Structure

1.1 The Report Structure

1.2 The Report Structure (Graded Quiz)

1.3 Lesson Summary

2. Final Assignment

2.1 Final Assignment (Graded Peer-graded Assignment)

2.2 Final Assignment (Review Your Peers Graded)

3. Final Exam

3.1 Final Exam (Graded Quiz 12. Question)

3.2 Digital Badge

Conclusion

References

Introduction

Data Science is about data gathering, analysis and decision-making. Data Science is about finding patterns in data, through analysis, and make future predictions. By using Data Science, companies can make, Since, then people working in data science have carved out a unique and distinct field for the work they do. This field is data science. In this course, we will meet some data science practitioners and we will get an overview of what data science is today. In this module, you will view the course syllabus to learn what will be taught in this course. You will hear from data science professionals to discover what data science is, what data scientists do, and what tools and algorithms data scientists use daily. Finally, you will complete a reading assignment to find out why data science is considered the sexiest job in the 21st century. 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 analysing 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 analyse 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.

- Better decisions (should we choose A or B)
- Predictive analysis (what will happen next?)
- Pattern discoveries (find pattern, or maybe hidden information in the data)

WEEK 1

CHAPTER 1

Course Syllabus

Defining Data Science and What Data Scientists Do

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

What Do Data Scientists Do?

- A day in the Life of a Data Scientist
- Old problems, new problems, Data Science solutions
- Data Science Topics and Algorithms
- What is the cloud?
- What Makes Someone a Data Scientist?

Data Science Topics

- Foundations of Big Data
- How Big Data is Driving Digital Transformation
- What is Hadoop?
- Data Science Skills & Big Data
- Data Scientists at New York University
- Data Mining
- Quiz: Data Mining

Deep Learning and Machine Learning

- What's the difference?
- Neural Networks and Deep Learning
- Applications of Machine Learning
- Regression
- Quiz: Regression

Data Science in Business

- Applications of Data Science
- How Data Science is Saving Lives
- How Should Companies Get Started in Data Science?
- Applications of Data Science
- The Final Deliverable
- Quiz: The Final Deliverable

Careers and Recruiting in Data Science

- How Can Someone Become a Data Scientist?
- Recruiting for Data Science
- Careers in Data Science
- High School Students and Data Science Careers

The Report Structure

- The Report Structure
- Quiz: The Report Structure
- Final Assignment

CHAPTER 1.1

Professional Certificate Career Support

Completing a Professional Certificate on Coursera unlocks access to a private Professional Certificate Alumni Resources community, which provides exclusive career support resources, including Step-by-step guide to ensure your success at every stage of your job search. 1 year of free access to Big Interview's expert video lessons, resume builder, and interactive interview practice tools (a \$79/month value). A network and support of fellow completers of Coursera Professional Certificates. A variety of special offers such as career coaching, webinars, and more. After completing your Professional Certificate, you'll get an email telling you how to access these career support resources.

CHAPTER 2

Chapter 2.1

Defining 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. 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.



Fig: 2.1

Chapter 2.2

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. While 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, analyse 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.

Chapter 2.3

The Many Paths to Data Science

I went through many different stages in my life where I wanted to be a singer and then a doctor. And then I realized that I was good at math. So, I chose an area that was focused on quantitative analysis. And from then I do think that I wanted to work with data. Not necessarily data science as it's known today. The first time that I had contact with data science, when I was my first year as a mechanical engineering. And strategic consulting firms, they use data science to make decisions. So, it was my first contact with data science. I had a complicated problem that I needed to solve, and the usual techniques that we had at that time couldn't help with that problem. I graduated with a math degree in the worst possible time, right after the economic crisis, and you had to be useful to get a job.

Chapter 2.4

Advice for New Data Scientists

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 let us 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

Chapter 2.5

The Sexiest Job in the 21st Century

In the data-driven world, data scientists have emerged as a hot commodity. The chase is on to find the best talent in data science. Already, experts estimate that millions of jobs in data science might remain vacant for the lack of readily available talent. The global search for skilled data scientists is not merely a search for statisticians or computer scientists. In fact, the firms are searching for well-rounded individuals who possess the subject matter expertise, some experience in software programming and analytics, and exceptional communication skills. Our digital footprint has expanded rapidly over the past 10 years. The size of the digital universe was roughly 130 billion gigabytes in 1995. By 2020, this number will swell to 40 trillion gigabytes. Companies will compete for hundreds of thousands, if not millions, of new workers needed to navigate the digital world. No wonder the prestigious Harvard Business Review called data science **the sexiest job in the 21st century**.

Chapter 2.6

The Sexiest Job in the 21st Century (Quiz)

Data Science: The Sexiest Job in the 21st Century
Graded Quiz • 9 min

Due Jun 20, 12:29 PM IST

✔ Congratulations! You passed!

Grade received 100%	Latest Submission Grade 100%	To pass 66% or higher	Go to next item
---------------------	------------------------------	-----------------------	-----------------

1. Walmart addressed its analytical needs by approaching Kaggle to host a competition for analyzing its proprietary data. 1 / 1 point

☐ False.

☒ True.

✔ Correct
Correct.

2. What is the average base salary of a data scientist reported by the New York Times? 1 / 1 point

☐ \$100,000

☐ \$150,000

☒ \$112,000

☐ \$16 per hour

☐ \$85,000 + Bonus

✔ Correct

Activate Windows
Go to Settings to activate Windows.

Fig: 2.2

Chapter 2.7

Lesson Summary

In this lesson, I have learned:

- Data science is the study of large quantities of data, which can reveal insights that help organizations make strategic choices.
- There are many paths to a career in data science; most, but not all, involve a little math, a little science, and a lot of curiosity about data.
- New data scientists need to be curious, judgemental, and argumentative.
- Why data science is considered the sexiest job in the 21st century, paying high salaries for skilled workers.

CHAPTER 3

What Do Data Scientists Do?

A data scientist's daily tasks revolve around data, which is no surprise given the job title. Data scientists spend much of their time gathering data, looking at data, shaping data, but in many ways and for many different reasons.

Chapter 3.1

A day in the Life of a Data Scientist

I've built a recommendation engine before, as part of a large organization and worked through all types of engineers and accounted for different parts of the problem. It's one of the ones I'm most happy with because ultimately, I came up with a very simple solution that was easy to understand from all levels, from the executives to the engineers and developers. Ultimately, it was just as efficient as something complex, and they could have spent a lot more time on. Back in the university, we have a problem that we wanted to predict algae blooms. This algae blooms could cause a rise in toxicity of the water, and it could cause problems through the water treatment company. We couldn't like to predict with our chemical engineering background. So, we use artificial neural

networks to predict when these blooms will reoccur. So, the water treatment companies could better handle this problem.

Chapter 3.2

Old problems, new problems, Data Science Solutions

Organizations can leverage the almost unlimited amount of data now available to them in a growing number of ways. However, all organizations ultimately use data science for the same reason—to discover optimum solutions to existing problems. Let's look at three examples of data science providing innovative solutions for old problems. In transport, Uber collects real-time user data to discover how many drivers are available, if more are needed, and if they should allow a surge charge to attract more drivers. Uber uses data to put the right number of drivers in the right place, at the right time,

for a cost the rider is willing to pay. In a different transport related data science effort, the Toronto Transportation Commission has made great strides in solving an old problem with traffic flows, restructuring those flows in and around the city. Using data science tools and analysis, they have: Gathered data to better understand streetcar operations, and identify areas for interventions Analysed customer complaints data, Used probe data to better understand traffic performance on main routes and created a team to better capitalize on big data for both planning, operations and evaluation By focusing on peak hour clearances and identifying the most congested routes, monthly hours lost for commuters due to traffic congestion dropped from 4.75 hrs. in 2010 to 3 hrs. in mid-2014.

Chapter 3.3

Data Science Topics and Algorithms

I really enjoy regression. I'd say regression was maybe one of the first concepts that I, that really helped me understand data, so I enjoy regression. I really like data visualization. I think it's a key element for people to get across their message to people that don't understand that well what data science is. Artificial neural networks. I'm passionate about neural networks because we have a lot to

learn with nature so when we are trying to mimic our, our brain I think that we can do some applications with this behaviour with this biological behaviour in algorithms. Data visualization with R. I love to do this. Nearest neighbour. It's the simplest but it just gets the best results so many more times than some overblown, overworked algorithm that's just as likely to overfit as it is to make a good fit. So structured data is more like tabular data things that you're familiar with in Microsoft Excel format. You've got rows and columns and that's called structured data.

Unstructured data is basically data that is coming from mostly from web where it's not tabular. It is not, it's not in rows and columns. It's text. It's sometimes it's video and audio, so you would have to deploy more sophisticated algorithms to extract data. And in fact, a lot of times we take unstructured data and spend a great deal of time and effort to get some structure out of it and then analyse it.

Chapter 3.4

Introduce Yourself

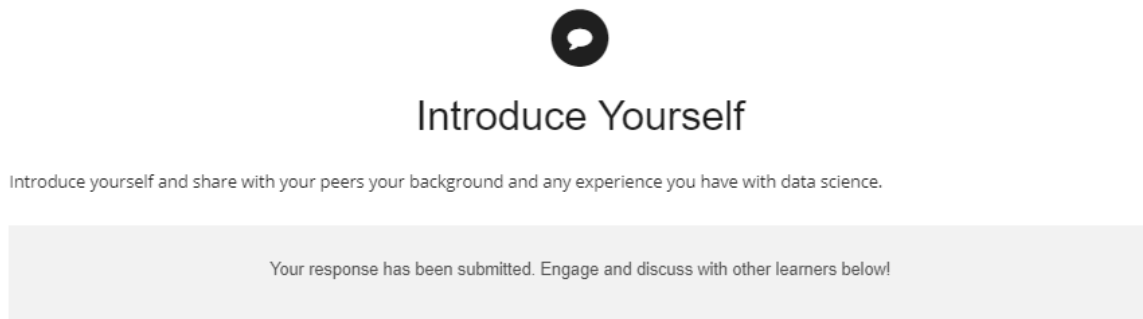


Fig: 3.1

Chapter 3.5

What Makes Someone a Data Scientist?

Now that you know what is in the book, it is time to put down some definitions. Despite their ubiquitous use, consensus evades the notions of Big Data and Data Science. The question Who is a data scientist? is very much alive and being contested by individuals, some of whom are merely

interested in protecting their discipline or academic turfs. In this section, I attempt to address these controversies and explain Why a narrowly construed definition of either Big Data or Data science will result in excluding hundreds of thousands of individuals who have recently turned to the emerging field.

Chapter 3.6

What Makes Someone a Data Scientist? (Quiz)

What Makes Someone a Data Scientist?
Graded Quiz • 9 min

✓ **Congratulations! You passed!**

Grade
received **100%**

Latest Submission
Grade **100%**

To pass 66% or
higher

[Go to next item](#)

1. According to the reading, how does **the author** define **data science**?

1 / 1 point

- ☐ Data science is some data and more science.
- ☒ Data science is what data scientists do.
- ☐ Data science is a way of understanding things and understanding the world.
- ☐ Data science is a physical science like physics or chemistry
- ☐ Data science is the art of uncovering the hidden secrets in data.

✓ **Correct**

True. This is the author's definition of data science.

2. According to the reading, what is admirable about **Dr. Patil's** definition of a **data scientist**?

1 / 1 point

- ☐ His definition excludes statistics.
- ☐ His definition limits data science to activities involving machine learning.
- ☒ His definition is inclusive of individuals from various academic backgrounds and training.
- ☐ His definition is about weaving strong narratives into analytics.

✓ **Correct**

Correct.

Fig: 3.2

Chapter 3.7

Lesson Summary

In this lesson, I have learned:

- The typical workday for a Data Scientist varies depending on what type of project they are working on.
- Many algorithms are used to bring out insights from data.
- Accessing algorithms, tools, and data through the Cloud enables Data Scientists to stay up-to-date and collaborate easily.

Chapter 3.8

Learning Outcomes

- Define data science and its importance in today's data driven world.
- List some 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.
- Articulate why data science is considered the most in-demand job in the 21st century.
- Describe what a typical day in the life of a data scientist looks like.
- Define some of the commonly used terms in data science.

WEEK 2

CHAPTER 1

Big Data and Data Mining

Before discussing data mining, it's necessary to answer the question of just what the term "big data" refers to. In short, big data is characterized by its size — it consists of datasets so large that they require the assistance of computer technology to be analyzed. According to Data Science Central, the term "big data" first emerged in 1997 and was used to refer to data collections that were too large to be "captured within an acceptable scope." In the decade that followed, the term was redefined several times. The concept as we understand it today was introduced to the wider public in 2007, according to the World Economic Forum. To qualify as big data as it's now commonly understood and accepted, the following criteria must be met, known as the five V's:

Volume. A very large amount of information is required — usually at least 1 terabyte of data.

Variety. Big data is further characterized by the fact that it comes from a wide variety of sources, such as social media, web servers, photos, and audio recordings.

Velocity. Big data is also set apart by fast growth; it must be increasing at a rapid, ideally exponential, rate.

Veracity. Veracity refers to how accurate or trustworthy the data is.

Value. Big data must have value. Data scientists should be able to use techniques like data mining to discern this value and yield a benefit for the companies they work for.

Without big data, data mining wouldn't exist. Data mining describes the process by which companies study information to gain insights into consumer behavior. Every modern industry relies on data mining in some way — and usually uses this information to improve consumers' lives. Data mining refers specifically to the process of finding meaning in expansive volumes of data. Data scientists collect large amounts of data and study it, looking for patterns and discrepancies to solve problems. Take the example of the grocery store in the introduction: Data can be automatically collected as customers swipe their loyalty cards, with their purchases noted, what day of the week they purchased

items, and even what time of day they made their purchases.

Preprogrammed algorithms sort purchases into an ordered Microsoft Excel table. The data scientist needs to examine this raw data, but it's not humanly possible to read through such a volume of information.

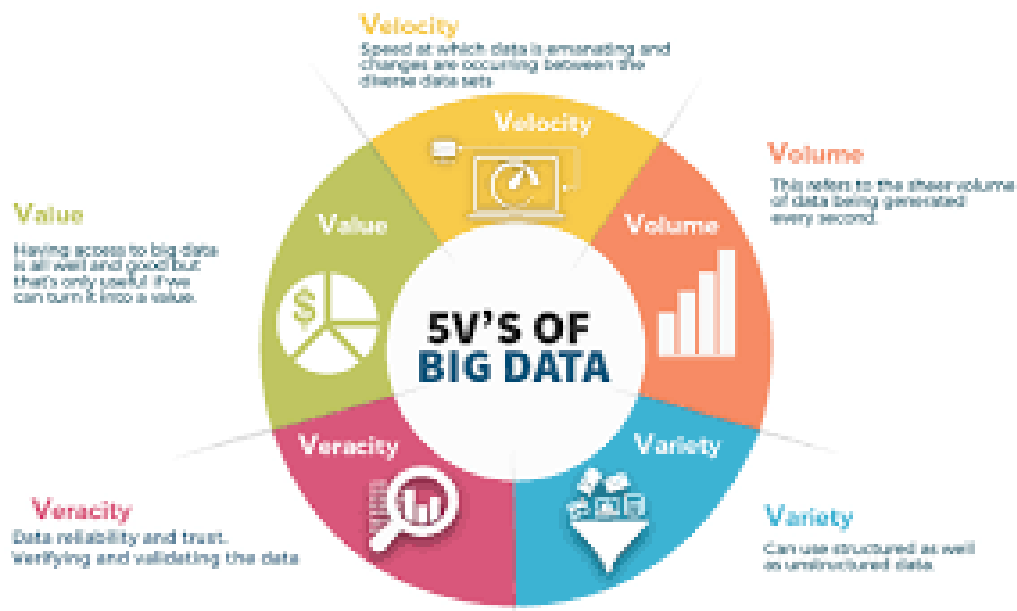


Fig: 1.1

Chapter 1.1

Foundation of Big Data

This badge earner has a basic understanding of Big Data concepts and their applications to gain insight for providing better service to customers. The earner understands that Big Data should be processed in a platform that can handle the variety, velocity, and the volume of data by using components that require integration and data governance.

Chapter 1.2

What is Hadoop?

Hadoop project develops open-source software for reliable, scalable, distributed computing.

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly available service on top of a cluster of computers, each of which may be prone to failures.

Chapter 1.3

How Big Data is Driving Digital Transformation

Digital Transformation affects business operations, updating existing processes and operations and creating new ones to harness the benefits of new technologies. This digital change integrates digital technology into all areas of an organization resulting in fundamental changes to how it operates and delivers value to customers. It is an organizational and cultural change driven by Data Science, and especially Big Data. The availability of vast amounts of data, and the competitive advantage that analyzing it brings, has triggered digital transformations throughout many industries. Netflix moved from being a postal DVD lending system to one of the world's foremost video streaming providers, the Houston Rockets NBA team used data gathered by overhead cameras to analyze the most productive plays, and Lufthansa analyzed customer data to improve its service. Organizations all around us are changing to their very core. Let's look at an example, to see how Big Data can trigger a digital transformation, not just in one organization, but in an entire industry. In 2018, the Houston Rockets, a National Basketball Association, or NBA team, raised their game using Big Data.

Chapter 1.4

Data Science Skills & Big Data

Programming:

You need to have knowledge of various programming languages, such as Python, Perl, C/C++, SQL, and Java, with Python being the most common coding language required in data science roles. These programming languages help data scientists organize unstructured data sets.

Knowledge of SAS and Other Analytical Tools

An understanding of analytical tools is a helpful data scientist skill for extracting valuable information from an organized data set. SAS, Hadoop, Spark, Hive, Pig, and R are the most popular data analytical tools that data scientists use. Certifications can help you establish your expertise in these analytical tools and help you gain this valuable data scientist skill!

Strong Communication Skills

Next on the list of top data scientist skills is communication. Data scientists clearly understand how to extract, understand, and analyse data. However, for you to be successful in your role, and for your organization to benefit from your services, you should be able to successfully communicate your findings with team members who don't have the same professional background as you.

BIG DATA-

Big data is a combination of structured, semi structured, and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modelling, and other advanced analytics applications. Systems that process and store big data have become a common component of data management architectures in organizations, combined with tools that support big data analytics uses. Big data is often characterized by the three V's:

Chapter 1.5

Data Mining

The first step in data mining requires you to set up goals for the exercise. Obviously, you must identify the key questions that need to be answered. However, going beyond identifying the key questions are the concerns about the costs and benefits of the exercise. Furthermore, you must determine, in advance, the expected level of accuracy and usefulness of the results obtained from data mining.

Selecting Data

The output of a data-mining exercise largely depends upon the quality of data being used. At times, data are readily available for further processing. For instance, retailers often possess large databases of customer purchases and demographics. On the other hand, data may not be readily available for data mining.

Pre-processing Data

Pre-processing data is an important step in data mining. Often raw data are messy, containing erroneous or irrelevant data. In addition, even with relevant data, information is sometimes missing. In the pre-processing stage, you identify the irrelevant attributes of data and expunge such attributes from further consideration. At the same time, identifying the erroneous aspects of the data set and flagging them as such is necessary. For instance, human error might lead to inadvertent merging or incorrect parsing of information between columns.

If the data were missing randomly, a simple set of solutions would suffice

Transforming Data

After the relevant attributes of data have been retained, the next step is to determine the appropriate format in which data must be stored. An important consideration in data mining is to reduce the number of attributes needed to explain the phenomena. This may require transforming data. Often you need to transform variables from one type to another. It may be prudent to transform the

continuous variable for income into a categorical variable where each record in the database is identified as low, medium, and high-income individual.

Storing Data

The transformed data must be stored in a format that makes it conducive for data mining. The data must be stored in a format that gives unrestricted and immediate read/write privileges to the data scientist. During data mining, new variables are created, which are written back to the original database, which is why the data storage scheme should facilitate efficiently reading from and writing to the database. Data safety and privacy should be a prime concern for storing data.

Mining Data

After data is appropriately processed, transformed, and stored, it is subject to data mining. This step covers data analysis methods, including parametric and non-parametric methods, and machine-learning algorithms. A good starting point for data mining is data visualization. Multidimensional views of the data using the advanced graphing capabilities of data mining software are very helpful in developing a preliminary understanding of the trends hidden in the data set.

Chapter 1.6

Data Mining (Graded Quiz)

Data Mining
Graded Quiz • 3 min

🎉 Congratulations! You passed!

Grade received 100%	Latest Submission Grade 100%	To pass 66% or higher	Go to next item
---------------------	------------------------------	-----------------------	---------------------------------

1. Prior Variable Analysis and Principal Component Analysis are both examples of a data reduction algorithm. 1 / 1 point

☐ True.

☒ False.

🟢 Correct
Correct. Prior Variable Analysis is not a data reduction algorithm.

2. After the data are appropriately processed, transformed, and stored, machine learning and non-parametric methods are a good starting point for data mining. 1 / 1 point

☒ False.

☐ True.

🟢 Correct
Correct.

3. "Formal evaluation could include testing the predictive capabilities of the models on observed data to see how effective and efficient the algorithms have been in reproducing data." This is known as: 1 / 1 point

☒ In-sample forecast.

☐ Overfitting.

☐ Reverse engineering.

Fig: 1.2

Chapter 1.7

Lesson Summary

In this lesson, I have learned:

- How Big Data is defined by the Vs: Velocity, Volume, Variety, Veracity, and Value.
- How Hadoop and other tools, combined with distributed computing power, are used to handle the demands of Big Data.
- What skills are required to analyse Big Data.
- About the process of Data Mining, and how it produces results.

CHAPTER 2

Deep Learning and Machine Learning

Deep Learning:

Deep Learning is a subset of Machine Learning where the artificial neural network and the recurrent neural network come in relation. The algorithms are created exactly just like machine learning, but it consists of many more levels of algorithms. All these networks of the algorithm are together called the artificial neural network. In much simpler terms, it replicates just like the human brain as all the neural networks are connected in the brain, which exactly is the concept of deep learning. It solves all the complex problems with the help of algorithms and its process.

Machine Learning:

Machine Learning: Machine learning is a subset, an application of Artificial Intelligence (AI) that offers the ability to the system to learn and improve from experience without being programmed to that level. Machine Learning uses data to train and find accurate results. Machine learning focuses on the development of a computer program that accesses the data and uses it to learn from itself.

Chapter 2.1

Neural Networks and Deep Learning

Neural networks, a beautiful biologically inspired programming paradigm which enables a computer to learn from observational data

Deep learning, a powerful set of techniques for learning in neural networks.

Neural networks and deep learning currently provide the best solutions to many problems in image recognition, speech recognition, and natural language processing. This book will teach you many of the core concepts behind neural networks and deep learning.

Chapter 2.2

Applications of Machine Learning

Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. Below are some most trending real-world applications of Machine Learning:

Image Recognition:

Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is Automatic friend tagging suggestion:

Speech Recognition

While using Google, we get an option of "Search by voice," it comes under speech recognition, and it's a popular application of machine learning.

Traffic prediction:

If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.

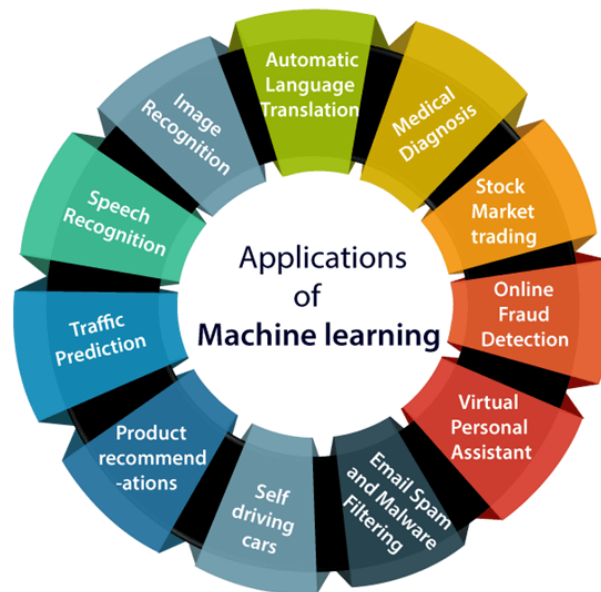


Fig: 2.1

Chapter 2.3

Regression

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

Regression analysis is a powerful tool for uncovering the associations between variables observed in data but cannot easily indicate causation. It is used in several contexts in business, finance, and economics. For instance, it is used to help investment managers value assets and understand the relationships between factors such as commodity prices and the stocks of businesses dealing in those commodities.

Chapter 2.4

Regression (Graded Quiz)

Regression
Graded Quiz • 9 min

✓ Congratulations! You passed!

Grade
received 100%

Latest Submission
Grade 100%

To pass 66% or
higher

Go to next item

1. Who developed the statistical technique known as regression?

1 / 1 point

- ☐ Thomas Bayes
- ☒ Sir Frances Galton
- ☐ Gerolamo Cardano
- ☐ Blaise Pascal
- ☐ Sir Isaac Newton

✓ Correct

2. The author discovered that houses located more than 2.5 kms to shopping centres sold for less than the rest.

1 / 1 point

- ☐ True.
- ☒ False.

✓ Correct

Correct. It is the other way around; houses located less than 2.5 kms to shopping centres sold for less than the rest.

Fig: 2.2

Chapter 2.5

Lesson Summary

In this lesson, I have learned:

The differences between some common Data Science terms, including Deep Learning and Machine Learning.

Deep Learning is a type of Machine Learning that simulates human decision-making using neural

networks.

Machine Learning has many applications, from recommender systems that provide relevant choices for customers on commercial websites, to detailed analysis of financial markets.

How to use regression to analyze data.

Chapter 3

Hands-on Exercise: Data Science Exploration

Chapter 3.1

Obtain an IBM Cloud Feature Code

The screenshot shows a course interface with a sidebar on the left and a main content area. The sidebar contains a breadcrumb trail: 'What is Data Science?' > 'Week 2' > 'Obtain an IBM Cloud Feature Code'. Below this, the sidebar lists course sections: 'Big Data and Data Mining', 'Deep Learning and Machine Learning', and 'Hands-on Exercise: Data Science Exploration'. Under the last section, there is a list of labs: 'Ungraded External Tool: Obtain an IBM Cloud Feature Code' (1h), 'Lab: Create IBM Cloud Account' (15 min), 'Lab: Create Watson Studio Instance' (15 min), and 'Lab: Exploring Data using Watson Studio & Jupyter' (15 min). The main content area has the title 'Obtain an IBM Cloud Feature Code'. It contains the following text: 'To facilitate hands-on skills development and enable you to complete some of the labs in this course, we are providing you with a special feature code for IBM Cloud.'; 'This feature code will enable you to either create an IBM Cloud account without a credit card or apply the feature code to your account, if you have one already created (and there are no feature code already applied to your account).'; 'Please check the box and click on the "Open Tool" button below to get a unique Feature Code to activate your trial.'; 'We have also provided you with the stepwise instructions on how to create an IBM Cloud account for your reference in the next section.'; and a note: 'NOTE: If you have already applied your IBM Cloud feature code in another course/lab to create an IBM Cloud account, please skip this item, as the code can only be used once.' Below this text, there is a horizontal line and then the text: 'This course uses a third-party tool, Obtain an IBM Cloud Feature Code, to enhance your learning experience. The tool will reference basic information like your name, email, and Coursera ID.' At the bottom, there is a 'Coursera Honor Code' section with a checked checkbox and the text 'I agree to use this tool responsibly.' and a blue 'Open Tool' button with an external link icon.

What is Data Science? > Week 2 > Obtain an IBM Cloud Feature Code

Big Data and Data Mining
Deep Learning and Machine Learning
Hands-on Exercise: Data Science Exploration

- Ungraded External Tool: Obtain an IBM Cloud Feature Code 1h
- Lab: Create IBM Cloud Account 15 min
- Lab: Create Watson Studio Instance 15 min
- Lab: Exploring Data using Watson Studio & Jupyter 15 min

Obtain an IBM Cloud Feature Code

To facilitate hands-on skills development and enable you to complete some of the labs in this course, we are providing you with a special feature code for IBM Cloud.

This feature code will enable you to either create an IBM Cloud account without a credit card or apply the feature code to your account, if you have one already created (and there are no feature code already applied to your account).

Please check the box and click on the "Open Tool" button below to get a unique Feature Code to activate your trial.

We have also provided you with the stepwise instructions on how to create an IBM Cloud account for your reference in the next section.

NOTE: If you have already applied your IBM Cloud feature code in another course/lab to create an IBM Cloud account, please skip this item, as the code can only be used once.

This course uses a third-party tool, Obtain an IBM Cloud Feature Code, to enhance your learning experience. The tool will reference basic information like your name, email, and Coursera ID.

Coursera Honor Code
☒ I agree to use this tool responsibly.

[Open Tool](#)

Fig: 3.1

Chapter 3.2

Create IBM Cloud Account

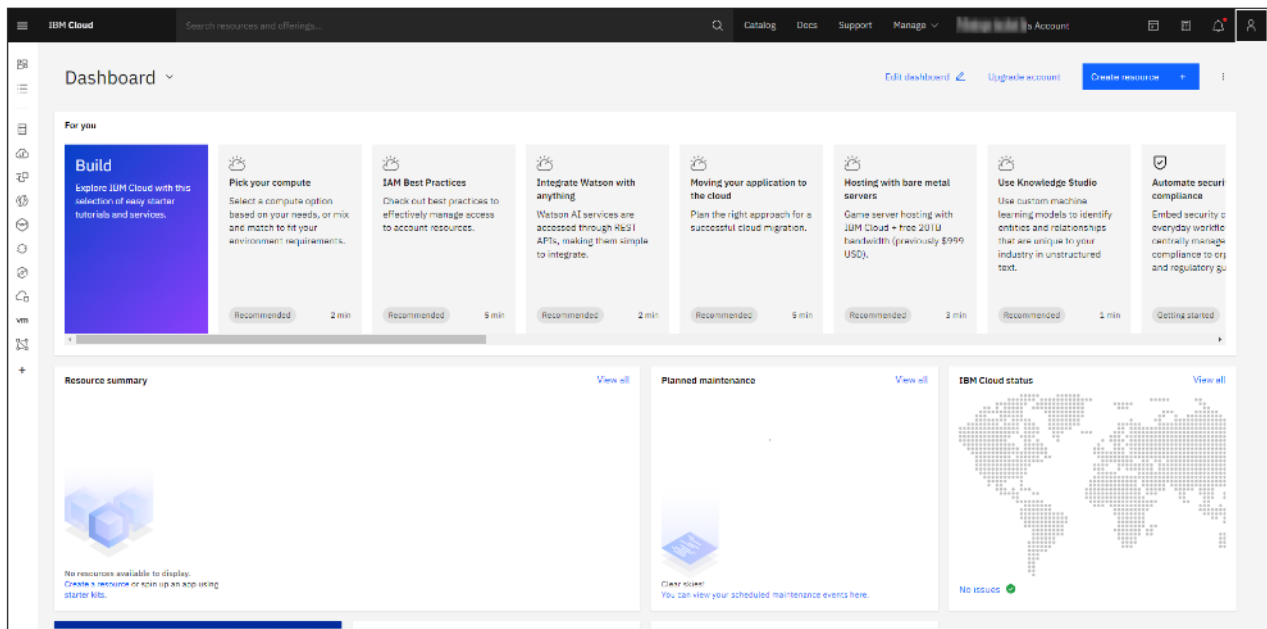


Fig: 3.2

Chapter 3.3

Lab: Create Watson Studio Instance

Objectives

Hands on Lab: Watson Studio Service creation

After completing this lab, you will be able to:

Use Watson Studio for your data science problem solving.

Create an instance of Watson Studio service:

- Go to IBM Cloud Login and login with your credentials.
- In the IBM Cloud Catalog search or choose AI/Machine Learning from Category and select Watson Studio.
- On the Watson Studio page, select an appropriate region depending on where you are accessing from, verify that the Lite plan is selected, and then click Create
- Once the Watson Studio instance is successfully created, click on Launch IBM Cloud Pak for Data.
- The screen below will appear for the first time you use Watson Studio.
- Once you get started, the first time, your IBM Cloud Pak for Data core services is provisioned.

Chapter 3.4

Learning Outcomes

1. Define Big Data and its distinguishing characteristics, such as velocity, volume, veracity, and value
2. Describe how Hadoop and other big data tools, combined with distributed computing power, are triggering digital transformation.
3. List some of the skills required to be a data scientist and analyse big data. List some of the skills required to analyze big data.
4. Explain what data mining is.
5. Summarize the importance of establishing goals, data selection, pre-processing, transformation, and storage of data in preparation for data mining.
6. Explain the difference between deep learning and machine learning.
7. Describe regression and how it might be used to predict market behavior and trend analysis.

WEEK 3

CHAPTER 1

Data Science in Business

In this module, you will learn about the approaches companies can take to start working with data science. You will learn about some of the qualities that differentiate data scientists from other professionals. You will also learn about analytics, storytelling, and the pivotal role data scientists play in creating an effective final deliverable. Finally, you will apply what you learned about data science by answering open-ended questions.

Data Science helps businesses monitor, manage, and collect performance measures to improve decision-making across the organization. Companies may use trend analysis to make critical decisions to improve consumer engagement, corporate performance, and boost revenue.

Chapter 1.1

How Data Science is Saving Lives

Using Data Science techniques to understand and analyze the large data sets available today has a huge impact on human lives. It can provide targeted information to help healthcare professionals give the best treatment to patients or help predict natural disasters so that people can prepare early, and much more besides.

In healthcare, data scientists use predictive analytics developed from data mining, data modelling, statistics, and machine learning to find the best options for patients. This type of predictive analytics examines all known factors for a disease, including gene markers, associated conditions, and environmental factors. It then recommends appropriate tests, suitable trials, and any suggested treatments. Every individual physician has their own store of knowledge gained from their studies, interests, and experiences.

Data science systems that use predictive analytics ensure that all physicians can also access the latest information about the disease, tests, and treatment plans, tailored to their specific patient. With this type of system, every physician has access to the same knowledge, and the best options can be consistently offered, improving patient outcomes.

Data science tools enable organizations to analyse vast quantities of data from widely different sources, and present that information in a way that allows data scientists to gain new knowledge, in some cases, saving hundreds of lives.

Chapter 1.2

How Should Companies Get Started in Data Science?

At the end of the day, for businesses, they know one thing, that if they are unable to measure something, they are unable to improve it. And if they are unable to measure their costs, they are unable to reduce them. If they're unable to measure their profits, they are unable to increase them. So, the first thing a company has to do is to start recording information, start capturing data, data about costs. And the differentiate it by labour costs and material cost, the cost to how much it cost to sell one product and the total cost. And then you look at the revenue, where's your revenue coming from? Is 80% of your revenue coming from 20% of your customers? Or is it the other way around? So, first thing first, start capturing data. Once you have data, then you can apply algorithms and analytics to it. So, the first thing to do would be to capture data. If you're not capturing it, start capturing it. If you're capturing it, archive it. Do not overwrite on your old data thinking you don't need it anymore. Data never gets old. Data is always relevant, even if it's 100 years old, 200 years old. It is relevant to you and your firm and your success. So, keep data, capture it, archive it, make sure nothing goes to waste. Make sure there's a consistency. So, someone 20 years later trying to understand, that data should be able to do so, so have proper documentation. Do it now. Put the best practices for data archiving in place the moment you start a business. And if you're already in business and you haven't done it, do it now. >> Start measuring things. Too many companies haven't measured things properly for a decade and, then they decide, they want data science. Data science inside a company is only going to be as valuable as the data collected.

Chapter 1.3

Applications of Data Science

Data science and big data are making an undeniable impact on businesses, changing day-to-day operations, financial analytics, and especially interactions with customers. It's clear that businesses can gain enormous value from the insights data science can provide. But sometimes it's hard to see exactly how. So, let's look at some examples. In this era of big data, almost everyone generates masses of data every day, often without being aware of it. This digital trace reveals the patterns our online lives. If you have ever searched for or bought a product on a site like Amazon, you'll notice that it starts making recommendations related to your search. This type of system known as a recommendation engine is a common application of data science. Companies like Amazon, Netflix, and Spotify use algorithms to make specific recommendations derived from customer preferences and historical behaviour.

Data Science in Gaming

Image Recognition

In Health Care

Search Engines

In Finance

Medicine and Drug Development

Chapter 1.4

The Final Deliverable

The ultimate purpose of analytics is to communicate findings to the concerned who might use these insights to formulate policy or strategy. Analytics summarize findings in tables and plots. The data scientist should then use the insights to build the narrative to communicate the findings. In academia, the final deliverable is in the form of essays and reports. Such deliverables are usually 1,000 to 7,000 words in length. In consulting and business, the final deliverable takes on several forms. It can be a

small document of fewer than 1,500 words illustrated with tables and plots, or it could be a comprehensive document comprising several hundred pages. Large consulting firms, such as McKinsey and Deloitte, routinely generate analytics-driven reports to communicate their findings and, in the process, establish their expertise in specific knowledge domains.

Let's review the "United States Economic Forecast", a publication by the Deloitte University Press. This document serves as a good example for a deliverable that builds narrative from data and analytics. The 24-page report focuses on the state of the U.S. economy as observed in December 2014. The report opens with a grabber highlighting the fact that contrary to popular perception, the economic and job growth has been quite robust in the United States. The report is not merely a statement of facts.

Chapter 1.5

The Final Deliverable (Quiz)

The Final Deliverable
Graded Quiz • 9 min

✓ **Congratulations! You passed!**

Grade received 100%	Latest Submission Grade 100%	To pass 66% or higher	Go to next item
-------------------------------	--	--------------------------	---------------------------------

1. The United States Economic Forecast is a publication by McKinsey University Press. 1 / 1 point

☐ True.

☒ False.

✓ **Correct**
Incorrect.

2. The report discussed in the reading successfully did the job of: 1 / 1 point

☐ Convincing the leadership team to act on an initiative.

☒ Using data and analytics to generate the likely economic scenarios.

☐ Calculating projections for the economy.

☐ Summarizing pages and pages of research.

✓ **Correct**
Correct.

Fig: 1.1

Chapter 1.6

Lesson Summary

In this lesson, I have learned:

- Data Science helps physicians provide the best treatment for their patients, and helps meteorologists predict the extent of local weather events and can even help predict natural disasters like earthquakes and tornadoes.
- That companies can start on their data science journey by capturing data. Once they have data, they can begin analysing it.
- Some ways that data is generated by consumers.
- How businesses like Netflix, Amazon, UPs, Google, and Apple use the data generated by their consumers and employees.
- The purpose of the final deliverable of a Data Science project is to communicate new information and insights from the data analysis to key decision-makers

CHAPTER 2

Careers and Recruiting in Data Science

Chapter 2.1

How can Someone Becomes a data Scientist

A real data scientist, the high-end data scientists, are mostly PhDs. They often come out of physics, out of statistics, they have a computer science background, they have to have a math background, they have to know about databases and statistics and probability and all that stuff.

However, if you're coming into a data science team, I think the first skills you need is you need to know how to program, at least have some computational thinking, so having taken a programming

course, you need to know some algebra, at least up to analytics, geometry, and hopefully some calculus, some basic probability, some basic statistics, I mean really have to understand the difference and different statistical distributions, and database. I mean, one of the easiest places to start is relational databases, which stores lots and lots of our data so people can first walk before they can run by at least understanding about computers and databases and how we store things and if you understand relational databases nowadays you can still, just with that understanding, use big data clusters as if they were just a big relational database. You don't have to really have understand the whole MapReduce programming model. But then, as you go further up in the field, then you have to know a lot of computer science theory and statistics, it's really, and probability, it's really the intersection of them that the high-end data scientists, the PhD data scientists work with.

I do a lot of self-learning. I think everybody these days, I mean, I learned about Hadoop all by myself, I read some articles, I watched some videos, I thought, I played, although I'm a builder, I'm a tinkerer, so if I want to figure out how to do something, I build it.

Chapter 2.2

Recruiting for Data Science

When the companies are hiring people for a data science team, maybe a data scientist or an analyst, or a chief data scientist, the tendency would be to find the person who has all the skills, that they know the domain-specific knowledge. They're excellent in analyzing structured and unstructured data. And they're great at presenting and they've got great storytelling skills. So, if you put all this together, you will realize you're looking for a unicorn. And your odds of finding a unicorn are pretty rare. Because you can teach analytics skills, anyone can learn analytics skills if they dedicate time and effort to it. But what really matters is who's passionate about the kind of business that you do. Someone could be a great data scientist in the retail environment, but they may not be that excited about working in IT related firms or working with gigabytes of weblogs. But if someone is excited about those weblogs, if someone is excited about health-related data then they would be able to contribute to your productivity much more so. And I would say if I'm looking for someone, if I have to put together a data science team, I would first look for curiosity. Is that person curious about things not just for data science but anything like, are they curious about why this room is painted a certain

way, why do the bookshelves have books, and what kinds of books? They have to have a certain degree of curiosity about everything that is in their vision, that they look at. The second thing is do they have a sense of humor because, you see, you have to have a lighthearted about it. If someone is too serious about it, they probably would take it too seriously, and would not be able to look at the lighter elements. The third thing I think, and I think the last thing that I would look for if I had to have a hierarchy, the last thing I would look for are technical skills. I would go through the social skills, curiosity, and sense of humor. The ability to tell a story. The ability to know that there is a story there. And then once all is there then I would say, well, can you do the technical side of it? And if there is some hope or some sign of some technical skills.

Chapter 2.3

Careers in Data Science

The emergence of Internet of things and advances in distributed computing have brought vast amounts of data and the technological capability to analyze it. Now that we can extract useful insights and new knowledge, we need to know how to shape that data to focus on what to do with it and what it can do for us. Enter data science. Companies like LinkedIn, Glassdoor, indeed, and Dice track employment trends which show a career in data science moving up the list of most promising jobs to become number one since 2016. It remains one of the top three career choices for 2020. Dice noted that job postings are from companies in a wide variety of industries, not just tech. Global Industry Analysts Incorporated predicts that the data science platform market will grow by \$314.8 billion US by 2025, driven by a compounded growth of 38.2%. McKinsey Global Institute warned of huge talent shortages for data and analytics by 2018. Forrester Research Analyst Brandon Purcell said, in January of 2019, the demand for data scientists will only grow as organizations increasingly rely on data-driven insights. We're now well into that period, and recruiters are finding it difficult to fill the growing need for talented data scientists. What motivates someone going into a data science? For one thing, data science applies to almost any discipline. So , if you have the aptitude and desire to work with data, enjoy coding, have no problem learning math and statistics, and you are a good storyteller, then you can certainly enter a data science field and excel. For most people, this means acquiring additional tools and skills and continuously learning about new tools and techniques in the field.

Chapter 2.4

High School Students and Data Science Careers

Learn how to program. Learn some math. Take a course in probability. Learn a little bit of statistics. And then, play. Build something, write something. I mean, when I say build, programming and building systems, building things isn't just physical, right? You can build computer systems, statistical systems, whatever. But once you try to do something, then you'll know what tools you need, right? And you'll say, "Oh, oh my god, what? "There's this expression there, "what does an inner product mean? "What's that? "How do I, oh, okay, I can learn that." And then when they get to college, they will have a big jump on many of the other college students. And so, when they get out of college, they'll have an even bigger jump, and then make a lot of money. And they'll be happy, too. This stuff is fun, right? It's fun. If you're in high school and you're considering data science, I would say get familiar with data bases, start learning SQL, start thinking about, you know, computer science, if that's interesting. If you have a computer science course in your school, you know, take it, and that's a good part of being a data scientist. Beyond that there are probably ways to foster your creativity, right, your curiosity. If you like detective games, that's kind of cool, right. And if you like treasure hunts or whatever, right, if you're into that stuff, I think you'll, and you get the opportunity to do that stuff, that will help you as a data scientist because it's a really a good way to kind of make sure that you can be curious as you go about your daily life.

It's kind of the same thing, but learning from data and we're going through an election season right now, there's a lot of stuff in the news about polls and survey results and that's a great way to start a conversation and talk about how do the people who ran the polls, how do they know, how can they predict what's going to happen in the election. So that's another cool way to start a conversation about data science.

When you have a connection to the problem, it suddenly becomes much easier to use math to help understand it, I found. And so, you know, knowing the people who will benefit from the math that you do I think is really cool.

Chapter 2.5

Lesson Summary

In this lesson, I have learned:

- Data Scientists need programming, mathematics, and database skills, many of which can be gained through self-learning.
- Companies recruiting for a Data Science team need to understand the variety of different roles Data Scientists can play, and look for soft skills like storytelling and relationship building as well as technical skills.
- High school students considering a career in Data Science should learn programming, math, databases, and, most importantly practice their skills.

WEEK 4

CHAPTER 1

The Report Structure

Before starting the analysis, think about the structure of the report. Will it be a brief report of five or fewer pages, or will it be a longer document running more than 100 pages in length? The structure of the report depends on the length of the document. A brief report is more to the point and presents a summary of key findings. A detailed report incrementally builds the argument and contains details about other relevant works, research methodology, data sources, and intermediate findings along with the main results.

I have reviewed reports by leading consultants including Deloitte and McKinsey. I found that the length of the reports varied depending largely on the purpose of the report. Brief reports were drafted as commentaries on current trends and developments that attracted public or media attention. Detailed and comprehensive reports offered a critical review of the subject matter with extensive data analysis and commentary. Often, detailed reports collected new data or interviewed industry experts to answer the research questions.

Chapter 1.1

The Report Structure

As a data scientist, you are expected to do thorough analysis with the appropriate data, deploying the appropriate tools. As a writer, you are responsible for communicating your findings to the readers. Transport Policy, a leading research publication in transportation planning, offers a checklist for authors interested in publishing with the journal. The checklist is a series of questions authors are expected to consider before submitting their manuscripts to the journal. I believe the checklist is useful for budding data scientists and, therefore, I have reproduced it verbatim for their benefit.

- Have you told readers, at the outset, what they might gain by reading your paper?
- Have you made the aim of your work clear?
- Have you explained the significance of your contribution?
- Have you set your work in the appropriate context by giving sufficient background (including a complete set of relevant references) to your work?
- Have you addressed the question of practicality and usefulness?

Chapter 1.2

The Report Structure (Graded Quiz)

The Report Structure
Graded Quiz - 9 min

✓ **Congratulations! You passed!**

Grade received 100%	Latest Submission Grade 100%	To pass 66% or higher	Go to next item
-------------------------------	---------------------------------	--------------------------	---------------------------------

1. The results section is where you craft your main arguments and present your conclusion. 1 / 1 point

☐ True
☒ False

✓ **Correct**
Correct. The results section is where you present the empirical findings.
2. The discussion section is where you introduce the research methods and data sources used for the analysis. 1 / 1 point

☒ False.
☐ True.

✓ **Correct**
Correct.
3. Adding a list of references and an acknowledgment section are examples of housekeeping, according to the author. 1 / 1 point

☒ True.
☐ False.

✓ **Correct**

Fig: 1.1

Chapter 1.3

Lesson Summary

In this lesson, I have learned:

- The length and content of the final report will vary depending on the needs of the project.
- The structure of the final report for a Data Science project should include a cover page, table of contents, executive summary, detailed contents, acknowledgements, references, and appendices.
- The report should present a thorough analysis of the data and communicate the project findings.

CHAPTER 2

Final Assignment (Graded Peer-graded Assignment)

In this Assignment, you will demonstrate your understanding of the videos and the readings by answering open-ended questions, defining data science and data scientist, and describing the different sections comprising a final deliverable of a data science project. Please note that this assignment is worth 10% of your final grade.

Straightforward enlistment process investigation

Submitted on June 30, 2022

[Shareable Link](#)

PROMPT	RUBRIC
<p>Based on the videos and the reading material, how would you define a data scientist and data science? (3 marks)</p> <p>A data scientist is a professional responsible for collection, examining and deciphering incredibly a lot of information. A scientist job is a branch-off of a few customary specialized jobs, including mathematician, scientist, analyst and PC proficient. A data scientist, in view of my momentum getting it, is the individual who draws a obvious conclusion regarding the business world and the information world. Likewise, information science is the art that an data scientist uses to get this going. Information science and AI gives the premise o business development, cost and chance decrease and, surprisingly, new plan of action creation, worked for data scientists by data scientists, now you can make esteem quicker utilizing the best of open source and IBM together. Information science is an itemized investigation of the progression of data from the goliath measures of information present is an associations vault. It includes getting significant bits of knowledge from crude and unstructure information which is handled through insightful, programming and business abilities.</p>	<p>According to the reading material, a data scientist is someone who finds solutions to problems by analyzing data using appropriate tools and then tells stories to communicate their findings to the relevant stakeholders. Data science is defined as what data scientists do. Is the student's definition close to what is defined in the course material?</p> <ul style="list-style-type: none"> <input type="radio"/> 1 point Poor. The student provided a non-coherent definition of data science and data scientist. The student did not mention anything about using tools to find solutions to problems and communicating their findings through story-telling. <input type="radio"/> 2 points Good. The student provided definitions close to the ones discussed in the course material but the definitions are still incomplete. For example, the student forgot to mention that in addition to using tools to find solutions to problems, a data scientist also communicates their findings through story-telling. Or, the student forgot to mention that a data scientist, in addition to communicating their findings to the relevant stakeholders, they use tools to find solutions to

Fig 2.1

PROMPT	RUBRIC
<p>As discussed in the videos and the reading material, data science can be applied to problems across different industries. Give a brief explanation describing what industry you are passionate about and would like to pursue a data science career in? (2 marks)</p> <p>My dream is marketing-like data driven approach in IT-enlistment and straightforward course of examination of computerized promoting to recruiting efficiency. I need to make a strong data scientist.</p>	<p>Has the student shared what industry they are passionate about?</p> <ul style="list-style-type: none"> <input type="radio"/> 0 points No, the student did not supply what industry they were passionate about <input type="radio"/> 1 point Yes, but the student did not offer any explanation on why they were passionate about the industry. <input type="radio"/> 2 points Yes and the student also gave an explanation on why they were passionate about the industry.

Fig 2.2

PROMPT	RUBRIC
<p>Based on the videos and the reading material, what are the ten main components of a report that would be delivered at the end of a data science project? (5 marks)</p> <ol style="list-style-type: none"> 1. Discussion section 2. Cover page 3. Introductory section 4. Appendices 5. Table of content 6. References 7. Conclusion 8. Literature review 9. Methodology section 10. Result section 	<p>According to the course material, a final deliverable in the form of a report, has the following 10 main components:</p> <ol style="list-style-type: none"> 1. Cover page 2. Table of contents 3. Executive Summary 4. Introductory section 5. Methodology section 6. Results section 7. Discussion section 8. Conclusion section 9. References 10. Acknowledgment <p><input type="radio"/> 2 points Incomplete answer. Not all the ten components were listed.</p> <p><input type="radio"/> 5 points Complete answer. The student listed all ten components of a data science report.</p>

Fig 2.3

Chapter 2.1

Final Assignment (Review Your Peers Graded)

✓ Peer-graded Assignment: Final Assignment

Reviews 4 complete

ⓘ It looks like this is your first peer-graded assignment. [Learn more](#) ✕

Links or files from other learners aren't verified for security by Coursera. If a file appears suspicious, use your preferred antivirus software before opening.

You've finished your peer reviews

Well done! You sent 4 peers feedback that will help them. If you have time, please review one or two more. Every review you do helps another peer complete the course!

Review fellow learners

Award points and give constructive feedback using the grading criteria provided (if any). Follow these tips to make the most of your learning experience:

- Offer specific ideas or suggestions that the learner can apply right away. Be encouraging and supportive.
- Use a translation tool if you have a preferred language that's different from the learner you're reviewing. Language usage is never a grading criteria.
- Review 4 or more assignment submissions to receive a grade. Help more learners complete this course by giving more reviews.

Fig: 2.4

CHAPTER 3

Final Exam 3.1

Final Exam (Graded Quiz 12. Question)

Final Exam
Graded Quiz • 36 min

✓ Congratulations! You passed!

Grade
received 100%

Latest Submission
Grade 100%

To pass 80% or
higher

Go to next item

1. According to the Module 1 reading, "The Sexiest Job in the 21st Century", a report published by the McKinsey Global Institute, it is projected that there would be a shortage of 140,000 – 190,000 people with what type of skills in the United States by 2018?

1 / 1 point

- ☐ Social skills
- ☐ Management skills
- ☐ Economic skills
- ☒ Deep analytical skills

✓ Correct

2. According to the Module 1 reading, "The Sexiest Job in the 21st Century", what magazine called Data Science the sexiest job of the 21st century?

1 / 1 point

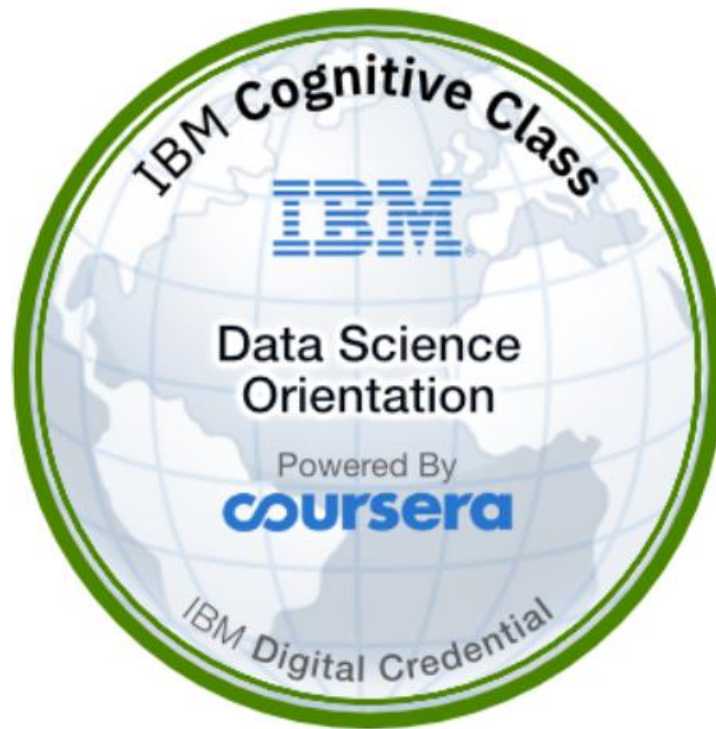
- ☐ Forbes
- ☐ Wired
- ☐ Bloomberg Businessweek
- ☒ Harvard Business Review

✓ Correct

Fig: 3.1

Chapter 3.2

IBM Digital Badge



This badge was issued to [Atul Kumar](#) on June 28, 2022

Conclusion

Data science is new, and we are still working out what it is. Now, it is best defined by what a data scientist does. A data scientist uses programming as the basis for a deeper, more flexible approach to data analysis. With its footprint in practically every industry, Data Science job-demand is estimated to rise several-fold in the future. The importance of Data Science is growing with every coming day. Data science is one of the growing fields. It has become an important part of almost every sector. It provides the best solutions that help to fulfill the challenges of the ever-increasing demand and maintainable future. As the importance of data science is increasing day by day, the need for a data scientist is also growing. Data scientists are the future of the world. Thus, a data scientist must be capable of providing great solutions which meet the challenges of all fields. To perform this, they should have proper resources and systems which help them to achieve their goal.

Data is one of the important features of every organization because it helps business leaders to make decisions based on facts, statistical numbers, and trends. Due to this growing scope of data, data science came into the picture which is a multidisciplinary field. It uses scientific approaches, procedures, algorithms, and frameworks to extract knowledge and insight from a huge amount of data. The extracted data can be either structured or unstructured. Data science is a concept to bring together ideas, data examination, Machine Learning, and their related strategies to comprehend and dissect genuine phenomena with data. Data science is an extension of various data analysis fields such as data mining, statistics, predictive analysis, and many more. Data Science is a huge field that uses a lot of methods and concepts which belong to other fields like information science, statistics, mathematics, and computer science. Some of the techniques utilized in Data Science encompass machine learning, visualization, pattern recognition, probability model, data engineering, signal processing, etc.

Bibliography

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0101EN-SkillsNetwork/labs/Module%202/Watson_Gallery.md.html?origin=www.coursera.org

https://en.wikipedia.org/wiki/Data_science

<https://www.javatpoint.com/life-cycle-phases-of-data-analytics>

<https://www.javatpoint.com/powerful-data-collection-tools-in-healthcare>

<https://www.geeksforgeeks.org/ml-introduction-to-transfer-learning/?ref=lbp>

<https://www.trifacta.com/hadoop-data-mining/>