Lesson 1

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, judgmental and argumentative.

Data science is considered the sexiest job in the 20th century, paying high salaries for skilled workers.

Lesson 2

The typical work day 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.

Lesson 3

In this lesson, you 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 analyze Big Data.

About the process of Data Mining, and how it produces results.

Lesson 4

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 analyzing 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.

Lesson 5

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.

Lesson 6

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.

In data science, there are many terms that are used interchangeably.

The term **big data** refers to data sets that are so massive, so quickly built, and so varied that they defy traditional analysis methods such as you might perform with a relational database.

The concurrent development of enormous compute power in distributed networks and new tools and techniques for data analysis means that organizations now have the power to analyze these vast data sets. A new knowledge and insights are becoming available to everyone. Big data is often described in terms of five V's;

velocity, volume, variety, veracity, and value.

**Data mining** is the process of automatically searching and analyzing data, discovering previously unrevealed patterns. It involves preprocessing the data to prepare it and transforming it into an appropriate format. Once this is done, insights and patterns are mined and extracted using various tools and techniques ranging from simple data visualization tools to machine learning and statistical models.

**Machine learning** is a subset of AI that uses computer algorithms to analyze data and make intelligent decisions based on what it is learned without being explicitly programmed. Machine learning algorithms are trained with large sets of data and they learn from examples. They do not follow rules-based algorithms. Machine learning is what enables machines to solve problems on their own and make accurate predictions using the provided data.

**Deep learning** is a specialized subset of machine learning that uses layered neural networks to simulate human decision-making. Deep learning algorithms can label and categorize information and identify patterns. It is what enables AI systems to continuously learn on the job and improve

the quality and accuracy of results by determining whether decisions were correct.

**Artificial neural networks**, often referred to simply as neural networks, take inspiration from biological neural networks, although they work quite a bit differently. A neural network in AI is a collection of small computing units called neurons that take incoming data and learn to make decisions over time. Neural networks are often layer-deep and are the reason deep learning algorithms become more efficient as the data sets increase in volume, as opposed to other machine learning algorithms that may plateau as data increases.

One more differentiation that is important to understand that between Artificial Intelligence and Data Science.

**Data Science** is the process and method for extracting knowledge and insights from large volumes of disparate data. It's an interdisciplinary field involving mathematics, statistical analysis, data visualization, machine learning, and more. It's what makes it possible for us to appropriate information, see patterns, find meaning from large volumes of data and use it to make decisions that drive business. Data Science can use many of the AI techniques to derive insight from data. For example, it could use machine learning algorithms and even deep learning models to extract meaning and draw inferences from data.

There is some **interaction between AI and Data Science**, but one is not a subset of the other. Rather, Data Science is a broad term that encompasses the entire data processing methodology while AI includes everything that allows computers to learn how to solve problems and make intelligent decisions.

Both AI and Data Science can involve the use of big data. That is, significantly large volumes of data.