Lessons Summary

In this lesson, you 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 20th century, paying high salaries for skilled workers.

**What Makes Someone a Data Scientist?**

**LATEST SUBMISSION GRADE**

80%

1.Question 1

Hal Varian, the chief economist at Google, declared that "the sexy job in the next ten years will be \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_".

**1 / 1 point**



Engineers.



Physicists.



Computer Scientists.



Statisticians.

**Correct**

Correct.

2.Question 2

**The author** defines a data scientist as someone who finds solutions to problems by analyzing data using appropriate tool and then tells stories to communicate their finding to the relevant stakeholders.

**1 / 1 point**



True.



False.

**Correct**

Correct.

3.Question 3

According to the reading, **the author** defines **data science** as the art of uncovering the hidden secrets in data.

**1 / 1 point**



True.



False.

**Correct**

Correct. The author defines data science as what data scientists do.

4.Question 4

What is admirable about **Dr. Patil’s** definition of a **data scientist** is that it limits data science to activities involving machine learning.

**1 / 1 point**



False.



True.

**Correct**

Correct. What is admirable about his definition is that it does not limit data science to activities involving machine learning.

5.Question 5

According to the reading, the characteristics exhibited by **the best** data scientists are those who are curious, ask good questions, and have at least 10 years of experience.

**0 / 1 point**



False.



True.

**Incorrect**

Incorrect.

**Data Mining**

**LATEST SUBMISSION GRADE**

80%

1.Question 1

According to the reading, the output of a data mining exercise largely depends on:

**1 / 1 point**



The data scientist.



The scope of the project.



The quality of the data.



The programming language used.

**Correct**

Correct.

2.Question 2

When data are missing in a systematic way, you can simply extrapolate the data or impute the missing data by filling in the average of the values around the missing data.

**1 / 1 point**



False.



True.

**Correct**

Correct. When data are missing in a systematic way, you should determine the impact of missing data on the results and whether missing data can be excluded from the analysis.

3.Question 3

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

**0 / 1 point**



True.



False.

**Incorrect**

Incorrect. Only Principal Component Analysis is a data reduction algorithm.

4.Question 4

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**



True.



False.

**Correct**

Correct.

5.Question 5

“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**



Overfitting.



Reverse engineering.



In-sample forecast.



Prototyping.

**Correct**

Correct.

According to the reading, the output of a data mining exercise largely depends on the skills of the data scientist carrying out the exercise.

**1 / 1 point**



True.



False.

**Correct**

Correct.

2.Question 2

What should you do when data are missing in a systematic way?

**1 / 1 point**



Extrapolate the data.



Determine who was managing the database.



Determine the average of the values around the missing data.



Determine the impact of missing data on the results and whether missing data can be excluded from the analysis.

**Correct**

Correct.

3.Question 3

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

**1 / 1 point**



False.



True.

**Correct**

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

4.Question 4

After the data are appropriately processed, transformed, and stored, what is a good starting point for data mining?

**1 / 1 point**



Creating a relational database.



Machine learning.



Data Visualization.



Non-parametric methods.

**Correct**

Correct.

5.Question 5

**In-sample forecast** is the process of formally evaluating the predictive capabilities of the models developed using observed data to see how effective the algorithms are in reproducing data.

**1 / 1 point**



True.



False.

**Correct**

Correct.

Neural networks have been around for decades, but due to religious reasons, people decided not to develop them any more because a neural network mimics the brain in the way it learns the data.



False.



True.

**This should not be selected**

Incorrect. They were abandoned for some time, not because of religious reasons, but because they were computationally very expensive.

Which of the following are use cases for deep learning?



Predicting the prices of houses using features such as number of bedrooms, square footage, and proximity to amenities.

**Un-selected is correct**

Predicting the prices of houses using features such as number of bedrooms, square footage, and proximity to amenities.

is not selected.This is correct.



Classifying images at a large scale.

**Correct**

Correct.

Classifying images at a large scale.

is selected.This is correct.

Correct.



Speech recognition.

**This should be selected**

**Regression**

**LATEST SUBMISSION GRADE**

100%

1.Question 1

The real added value of the author's research on residential real estate properties is quantifying the magnitude of relationships between housing prices and different determinants.

**1 / 1 point**



True.



False.

**Correct**

Correct.

2.Question 2

Regression is a statistical technique developed by Sir Frances Galton.

**1 / 1 point**



True.



False.

**Correct**

3.Question 3

According to the reading, the author discovered that an additional washroom adds more to the housing prices than an additional bedroom.

**1 / 1 point**



False.



True.

**Correct**

Correct.

4.Question 4

The author discovered that, *all else being equal*, houses located less than 5 kms but more than 2.5 kms to shopping centres sold for more than the rest.

**1 / 1 point**



True.



False.

**Correct**

Correct.

5.Question 5

"How much does a finished basement contribute to the price of a housing unit?" is a question that can be put to regression analysis.

**1 / 1 point**



False.



True.

**Correct**

Correct.