1. Data science is the field of exploring, manipulating, and analyzing data, and using data to answer questions or make recommendations.

True / False

1. Which of the following statements is correct?

As data science is not a discipline traditionally taught at universities, contemporary data scientists come from diverse backgrounds such as engineering, statistics, and physics.

To be a successful data scientist, it is a must to hold a degree in statistics or engineering.

This is correct. Contemporary data scientists come from different backgrounds such as engineering, mathematics, and even psychology. The secret skill is passion for continuous learning of new tools and patience to clean and analyze data.

1. According to professor Haider, the three important qualities to possess in order to succeed as a data scientist are:

Curious.

Judgemental

Argumentative.

Proficient in Programming.

Good at Math and Statistics.

Quiz 1

**1.**

Question 1

How is Walmart reported to have addressed its analytical needs?

1 point



Code sharing



Social media



None of the options is correct



Outsourcing



Crowdsourcing

**2.**

Question 2

The New York Times reported that the average base salary of a data scientist is $85,000 + competitive bonus.

1 point



False.



True.

**3.**

Question 3

According to professor Haider, the three important qualities to possess in order to succeed as a data scientist are:

1 point



Curious.



Good at Math and Statistics.



Good Story Teller (Argumentative).



Proficient in Programming.



Judgemental.

1. As a data scientist. You get to work on a broad spectrum of projects. Which of the following were projects discussed in the video?

Building a recommendation engine.

Correct. The engine was built by Stephen Sherman.

Analyzing hundreds of thousands of complaints filed with the Toronto Transit Commission.

Correct. This project was executed by professor Murtaza Haider.

Using neural networks to build a model that can predict falls in homes.

Building a classifier that can label fruits using shape and color as features.

1. Using complicated machine learning algorithms does not always guarantee achieving a better performance. Occasionally, a simple algorithm such as k-nearest neighbor can yield a satisfactory performance comparable to the one achieved using a complicated algorithm. It all depends on the data.

True / False.

In any field, and data science is no different, a simple solution is always preferred over a complicated one, especially if the performance is comparable.

1. According to professor Haidar, what is true about the cloud?

It allows you to bypass the physical limitations of your personal computer and the systems you are using.

One limitation of the cloud is that you are not able to deploy capabilities of advanced machines that do not necessarily have to be your machines.

Quiz 2

Question 1

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

Data science is the art of uncovering the hidden secrets in data.

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 some data and more science.

Data science is what data scientists do.

Question 2

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

False / True.

Question 3

According to the reading, what characteristics are said to be exhibited by the best data scientists?

Thinkers who are really curious and hold a Ph.D.

Curious individuals who ask good questions and are O.K. dealing with unstructured situations.

Really curious people who ask good questions.

Really curious engineers and statisticians.

Really curious people who ask good questions and have at least 10 years of experience.

WEEK 2

1. According to Dr. White, most of the components of data science, such as probability, statistics, linear algebra, and programming, have been around for many decades but now we have the computational capabilities to apply combine them and come up with new techniques and learning algorithms.

True / False

1. According to Dr. White, his students, who are mostly aspiring data scientists, need to learn many tools such as Python, UNIX commands, pandas, and Jupyter notebook.

TRUE / FALSE

1. Data science and business analytics have always been very hot subjects.

TRUE / FALSE

1. According to Dr. White, which of the following statements are correct about big data.

Big data is any data that is needs more than 10GB of memory to be stored.

Big data was started by Google when Google tried to figure out how how to solve their PageRank algorithm.

Big data is is data that is large enough and has enough volume and velocity that you cannot handle it with traditional data database systems.

1. Question 1
   1. What is an example of a data reduction algorithm?
   2. Prior Variable Analysis.
   3. Cojoint Analysis.
   4. A/B Testing.
   5. Principal Component Analysis.
2. After the data are appropriately processed, transformed, and stored, what is a good starting point for data mining?
   1. Machine learning.
   2. Non-parametric methods.
   3. Creating a relational database.
   4. Data Visualization.
3. 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.

TRUE / FALSE

1. 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.

TRUE / FALSE

1. Which of the following are use cases for deep learning?
   1. Predicting the prices of houses using features such as number of bedrooms, square footage, and proximity to amenities.
   2. Classifying images at a large scale.
   3. Speech recognition.
2. Netflix uses machine learning to recommend movies to you based on movies that you have already watched and liked or disliked.

TRUE / FALSE

1. Who developed the statistical technique known as regression?
2. Thomas Bayes
3. Sir Frances Galton
4. Sir Isaac Newton
5. Blaise Pasca
6. Gerolamo Cardan
7. What did the author's research reveal about proximity to large shopping centres?
8. The author discovered that houses located more than 5 kms to shopping centres sold for less than the rest.
9. The author discovered that proximity to large shopping centres had a nonlinear impact on the housing prices.
10. The author discovered that proximity to large shopping centres didn't have any significant impact on the prices of housing units.
11. The author discovered that houses located more than 2.5 kms to shopping centres sold for less than the rest.
12. "What are typical land taxes in a house sale?" is a question that can be put to regression analysis.

TRUE / FALSE

WEEK 3

1. What are some of the first steps that companies need to take to get started in data science?
2. Start Collecting data.
3. Put together a team of data scientists
4. Discard any old data that had acquired in order to start over.

Quiz 1

1. The Untied States Economic Forecast is a publication by McKinsey University Press.

T/F

1. The report discussed in the reading successfully did the job of using data and analytics to generate the likely economic scenarios.

T/ F

1. According to the reading, in order to produce a compelling narrative, initial planning and conceptualizing of the final deliverable is of extreme importance.

T/F

According to Dr. White, if someone is coming into a data science team, the first skills they would need are:



Knowing some algebra and some calculus.



Understanding relational databases.



Knowing how to program, at least have some computational thinking.



Knowing basic probability and some basic statistics.

1. Curiosity is one of the most important skills that a data scientist should possess.

T/F

1. When hiring a data scientist, you need to ensure that the candidate is passionate about your field of work.

T / F

1. According to Dr. White, the industrial world is shifting to a new trend, and for high school students to be on the right side of this new trend, his advice to them is:
   1. Take a course in probability.
   2. Learn some math.
   3. Learn statistics.
   4. Try to start experimenting with building small systems that work and are useful.
   5. To learn how to program.
2. The results section is where you craft your main arguments and present your conclusion.

T / F

1. The discussion section is where you:
   1. Highlight how your findings provide the ultimate missing piece to the puzzle.
   2. Introduce the research methods and data sources used for the analysis.
   3. Rely on the power of narrative to enable numbers to communicate your important findings to the readers.
   4. Refer the reader to the research question and the knowledge gaps you identified earlier.
2. According to the reading, what is an example of housekeeping?
   1. Saving the report as a PDF file.
   2. Adding headings to charts.
   3. Adding slide numbers.
   4. Adding a list of references.