- Welcome!
- About this course
- Module 1 - From Problem to Approach
- Module 2 - From Requirements to Collection
- Module 3 - From Understanding to Preparation
- Module 4 - From Modeling to Evaluation
- Module 5 - From Deployment to Feedback
- Course Summary
- **Final Exam**
  - **Instructions**
  - **Final Exam** Timed Exam
  - **Retake Exam**
- Course Survey & Feedback
- Completion Certificate

Final Exam Instructions

1. Time allowed: **1 hour**

2. Attempts per question:

- One attempt - For True/False questions.
- Two attempts - For any question other than True/False.

3. Clicking the "**Final Check**" button when it appears, means your submission is **FINAL**. You will **NOT** be able to resubmit your answer for that question ever again.

IMPORTANT: Do not let the time run out and expect the system to grade you automatically. You must explicitly submit your answers, otherwise they would be marked as incomplete.

QUESTION 1 (1/1 point)

Select the correct sentence about the data science methodology explained in the course.

- ○ Data science methodology is not an iterative process – one does not go back and forth between methodological steps.
- ○ Data science methodology is a specific strategy that guides processes and activities relating to data science only for text analytics.
- ○ Data science methodology always starts with data collection.
- ◉ Data science methodology provides the data scientist with a framework for how to proceed to obtain answers. ✔
- ○ Data science methodology depends on a specific set of technologies or tools.

You have used 2 of 2 submissions

QUESTION 2 (1/1 point)

Business understanding is important in the data science methodology stage. Why?

- ○ Because it shapes the rest of the methodological steps.

Because it clearly defines the problem and the needs from a business perspective.

Because it ensures that the work generates the intended solution.

Because it involves domain expertise.

All of the above. ✔

You have used 2 of 2 submissions

QUESTION 3 (1/1 point)

A data scientist determines that building a recommender system is the solution for a particular business problem at hand. What stage of the data science methodology does this represent?

Modeling

Deployment

Model evaluation

Analytic approach ✔

Data understanding

You have used 2 of 2 submissions

QUESTION 4 (1/1 point)

Which of the following represent the two important characteristics of the data science methodology?

It is a highly iterative process and immediately ends when the model is deployed.

It is not an iterative process and it never ends.

- It immediately ends when the model is deployed because no feedback is required.
- It is a highly iterative process and it never ends. ✔

You have used 2 of 2 submissions

QUESTION 5 (1/1 point)

What do data scientists typically use for exploratory analysis of data and to get acquainted with them?

- They use support vector machines and neural networks as feature extraction techniques.
- They begin with regression, classification, or clustering.
- They use deep learning.
- They use descriptive statistics and data visualization techniques. ✔
- All of the above.

You have used 2 of 2 submissions

QUESTION 6 (1/1 point)

Select the correct statement about data preparation.

- Data preparation cannot be accelerated through automation.
- Data preparation involves dealing with missing improperly coded data and can include using text analysis to structure unstructured or semi-structured text data. ✔
- Data preparation is typically the least time-consuming methodological step.

- None of the above.

You have used 2 of 2 submissions

QUESTION 7 (1/1 point)

Which statement best describes the modeling stage of the data science methodology.

- Modeling is followed by the analytic approach stage.
- Modeling may require testing multiple algorithms and parameters. ✔
- Modeling is always based on predictive models.
- Modeling always uses training and test sets.
- All of the above.

You have used 2 of 2 submissions

QUESTION 8 (1/1 point)

Which of the following statements best describe the model evaluation stage of the data science methodology?

- Model evaluation may entail statistical significance tests, particularly when additional proof is necessary to justify some of the emerging recommendations.
- Model evaluation is important because it examines how well the model performs in the context of the business problem.
- Model evaluation entails computing graphs and/or various diagnostic measures such as a confusion matrix.
- Model evaluation is done using a test set if the model is a predictive one.
- All of the above. ✔

You have used 2 of 2 submissions

QUESTION 9 (1/1 point)

What does deploying a model into production represent?

- It represents the end of the iterative process that includes feedback, model refinement, and redeployment.
- It represents the beginning of an iterative process that includes feedback, model refinement and redeployment and requires the input of additional groups, such as marketing personnel and business owners. ✔
- It represents the final data science product.
- None of the above.

You have used 2 of 2 submissions

QUESTION 10 (1/1 point)

A data scientist, John, was asked to help reduce readmission rates at a local hospital. After some time, John provided a model that predicted which patients were more likely to be readmitted to the hospital and declared that his work was done. Which of the following best describes this scenario?

- John only provided one model as a solution and he should have provided multiple models.
- The scenario is already optimal.
- Even though John only submitted one solution, it might be a good one. However, John needed feedback on his model from the hospital to confirm that his model was able to address the problem appropriately and sufficiently. ✔
- John's mistake is that he lied in the *analytic approach* step of the data science methodology.
- John still needed to collect more data.

QUESTION 11 (1/1 point)

A car company asked a data scientist to determine what type of customers are more likely to purchase their vehicles. However, the data comes from several sources and is in a relatively “raw format”. What kind of processing can the data scientist perform on the data to prepare it for modeling?

- Feature engineering.
- Transforming the data into more useful variables.
- Combining the data from the various sources.
- Addressing missing/invalid values.
- All of the above. ✔

You have used 2 of 2 submissions

QUESTION 12 (1/1 point)

High-performance, massively parallel systems can be used to facilitate the following methodological steps.

- Data preparation and Modeling. ✔
- Modeling only.
- Deployment.
- Business understanding.
- All of the above.

You have used 2 of 2 submissions

QUESTION 13 (1/1 point)

- "Top-down" approach – the data, when sorted, is modeled from the "top" of the data towards the "bottom". "Bottom-up" approach – the data is modeled from the "bottom" of the data to the "top".
- "Top-down" approach – models are fit before the data is explored. "Bottom-up" approach – data is explored, and then a model is fit.
- (selected) "Top-down" approach – first defining a business problem then analyzing the data to find a solution. "Bottom-up" approach – starting with the data, and then coming up with a business problem based on the data. ✔
- "Top-down" approach – using massively parallel, warehouses with huge data volumes as the data source. "Bottom-up" approach – using a sample of small data before using large data.
- All of the above.

You have used 2 of 2 submissions

QUESTION 14 (1 point possible)

The following are all examples of rapidly evolving technologies that affect data science methodology **EXCEPT** for?

- (selected) Data sampling.
- Automation.
- Text analysis.
- Platform growth.
- In-database analytics.

You have used 2 of 2 submissions

QUESTION 15 (1/1 point)

- Discovering initial insights about the data.
- Visualizing the data.
- Assessing data quality.
- Understanding the content of the data.
- Gathering and analyzing feedback for assessment of the model's performance. ✔

You have used 2 of 2 submissions

QUESTION 16 (1/1 point)

For predictive models, a test set, which is similar to – but independent of – the training set, is used to determine how well the model predicts outcomes. This is an example of what step in the methodology?

- Data preparation.
- Deployment.
- Analytic approach.
- Model evaluation. ✔
- Data requirements.

You have used 2 of 2 submissions

QUESTION 17 (1/1 point)

"When _____ data is available (such as customer call center logs or physicians' notes in unstructured or semi-structured format), _____ analytics can be useful in deriving new structured variables to enrich the set predictors and improve model accuracy." Which of the following most appropriately fills in the blanks?

- market; statistical
- big; digital
- highly structured; text
- text; predictive

You have used 2 of 2 submissions

QUESTION 18 (1/1 point)

Typically in a predictive model, the training set and the test set are very different and independent, such as having a different set of variables or structure.

- True
- False ✔

You have used 1 of 1 submissions

QUESTION 19 (1/1 point)

Data scientists may frequently return to a previous stage to make adjustments, as they learn more about the data and the modeling.

- True ✔
- False

You have used 1 of 1 submissions

QUESTION 20 (1/1 point)

Why should data scientists maintain continuous communication with business sponsors throughout a project?

So that business sponsors can ensure the work remains on track to generate the intended solution.

So that business sponsors can review intermediate findings.

All of the above. ✔

None of the above.

You have used 2 of 2 submissions