

## COURSE 2- TOOLS FOR DATA SCIENCE

### MODULE 1

#### GRADED QUIZ 1

1. Which Data Science category enables you to present data in the form of charts, plots, and maps?

1 / 1 point

- ☐ Model Building
- ☐ Data Integration and Transformation
- ☐ Data Management
- ☒ Data Visualization

✓ Correct

Correct! Data visualization is the graphical representation of data and information. You can use visualization to represent data in charts, plots, maps, animations, and so forth.

2. Which Data Science category uses tools like Fiddler to track the performance of deployed models in a production environment?

1 / 1 point

- ☐ Data Visualization
- ☐ Data Integration and Transformation
- ☒ Model Monitoring

✓ Correct

Correct. Model monitoring uses tools like Fiddler to track the performance of deployed models in a production environment.

3. Which Data Science category enables you to train the data and analyze patterns using suitable machine learning algorithms?

1 / 1 point

- ☐ Model Monitoring and Assessment
- ☐ Model Deployment
- ☒ Model Building
- ☐ Data Visualization

✓ Correct

Correct. Model building is the step where you train the data and analyze patterns with machine learning algorithms.

4. Which of the following is a good example of code asset management?

1 / 1 point

- ☐ Watson Studio
- ☐ Apache Spark
- ☐ Jupyter Notebook
- ☒ GitHub

✓ **Correct**  
Correct. GitHub is a good example of code asset management.

5. What is the purpose of using Apache airflow?

1 / 1 point

- ☐ Data Management
- ☐ Data Visualization
- ☐ Model Deployment
- ☒ Data Integration and Transformation

✓ **Correct**  
Correct. The most widely used open-source data integration and transformation tools are the following: Apache AirFlow, which Airbnb originally created.

6. Which of the following statement is incorrect?

1 / 1 point

- ☐ Jupyter Notebooks unifies documentation, code, output from the code, shell commands, and visualizations in a single document
- ☐ JupyterLab is the next version of Jupyter Notebooks, and in the long-term will replace Jupyter Notebooks
- ☒ Jupyter Notebook is not an open source tool.
- ☐ Jupyter supports more than a hundred different programming languages through "kernels."

✓ **Correct**  
Correct. Review the Open Source Tools for Data Science - Part 2 video.

7. Which plot is used to plot a series of data points over time?

1 / 1 point

- ☐ Bar Chart
- ☐ Map Chart
- ☐ Tree Map
- ☒ Line Chart

✓ Correct

Correct. Various forms of data visualizations include a bar chart, which compares the size of each component, a treemap, which displays hierarchy data, a line chart, which plots a series of data points over time, and a map chart, which displays data by location.

8. Which of the following is the other name for Data Asset Management?

1 / 1 point

- ☐ Data Lakes
- ☐ Data Store
- ☒ Data governance
- ☐ Data Currency

✓ Correct

Correct. Data asset management is also known as data governance or data lineage.

9. What does model deployment use to enable data-based decisions?

1 / 1 point

- ☐ Packages
- ☒ APIs
- ☐ Gitlab
- ☐ Web Applications

✓ Correct

Correct. Model deployment uses APIs to enable data-based decisions.

10. Which of the following tool is a part of the IBM Watson Studio?

1 / 1 point

- ☐ Cloudant
- ☒ Data Refinery
- ☐ DataMeer
- ☐ Informatica

✓ Correct

Correct. Data Refinery is part of IBM Watson Studio.

## WEEK 2/ MODULE 2

### GRADED QUIZ

#### ✓ Congratulations! You passed!

Grade  
received **100%**

Latest Submission  
Grade **100%**

To pass 70% or  
higher

Retake the  
assignment in 2h  
31m

Go to  
next  
item

1. What does Python use for Natural Language Processing?

1 / 1 point

- ☐ Pandas
- ☐ Keras
- ☐ TensorFlow
- ☒ Natural Language Toolkit

✓ **Correct**

Correct! Python can also be used for Natural Language Processing (NLP) using the Natural Language Toolkit (NLTK).

2. In Python, which library is not used for Data Science?

1 / 1 point

- ☐ Pandas
- ☐ SciPy
- ☒ Caffe
- ☐ Natural Language Toolkit

✓ **Correct**

Correct. Caffe is a deep learning framework made with expression, speed, and modularity in mind.

3. Which of the following is developed using the R language?

1 / 1 point

- ☐ Automation
- ☐ Relational DBMS
- ☒ Graphing
- ☐ Image processing

✓ **Correct**

Correct! R language is used for graphing.

4. Which statement is true for Open source and free software?

1 / 1 point

- ☒ Open source is more business focused, while free software is more focused on a set of values.
- ☐ Both open source and free software doesn't support collaboration.
- ☐ Open source is more focused on a set of values, while free software is more business focused.
- ☐ None of the above.

✓ **Correct**

Correct. The Open-Source Initiative (OSI) champions open source, while the Free Software Foundation (FSF) defines free software. Open source is more business focused, while free software is more focused on a set of values.

5. Which of the following is a global R community?

1 / 1 point

- ☒ useR
- ☐ RealPython
- ☐ HackerEarth
- ☐ PyLadies

✓ **Correct**

Correct! useR is an R community.

6. Is the following statement true or false: R integrates well with C and C++..

1 / 1 point

- ☒ True
- ☐ False

✓ **Correct**  
R integrates well with other computer languages like C++, Java, C, .Net and Python.

7. Which of the following is not an SQL language element?

1 / 1 point

- ☒ Spreadsheets
- ☐ Predicate
- ☐ Clauses
- ☐ Queries

✓ **Correct**  
Correct. Spreadsheet is not language but it is a tool.

---

8. Why is SQL an American National Standards Institute (or AN-see) standard?

1 / 1 point

- ☐ The SQL language is subdivided into several language elements
- ☒ SQL knowledge applies to multiple databases
- ☐ SQL can considerably speed up workflow executions
- ☐ SQL manages data

✓ **Correct**  
Correct! SQL is an American National Standards Institute (or AN-see) standard, which means if you learn SQL and use it with one database, you can apply your SQL knowledge to many other databases.

9. Which Scala program is used for data science?

1 / 1 point

- ☒ Apache Spark
- ☐ Caffe
- ☐ Weka
- ☐ MongoDB

✓ **Correct**  
Correct! The most popular program built with Scala for data science is Apache Spark.

10. Which is the language used to develop the Hadoop framework?

1 / 1 point

- ☐ Ruby
- ☐ Python
- ☒ Java
- ☐ C++

✓ **Correct**  
Hadoop is another application of Java which manages data processing and storage for big data applications running in clustered systems.

## WEEK 3/ MODULE 3

### GRADED QUIZ

✔ **Congratulations! You passed!**

Grade  
received 100%

Latest Submission  
Grade 100%

To pass 70% or  
higher

Go to next item

1. Which library is used for machine learning?

1 / 1 point

- ☒ Scikit-learn
- ☐ Matplotlib
- ☐ Pandas
- ☐ NumPy

✔ **Correct**

Correct. For machine learning, the Scikit-learn library contains tools for statistical modeling, including regression, classification, clustering, and so on.

2. What is the main purpose of an Application Programming Interface (API)?

1 / 1 point

- ☐ Facilitate communication between hardware components
- ☐ Create graphical representations of data
- ☒ Allow communication between software components
- ☐ Design user interfaces for applications

✔ **Correct**

Correct. An API allows communication between software components.

3. In a REST API architecture, what does the client typically receive from the web service after sending a request?

1 / 1 point

- ☐ HTML file
- ☐ Binary data
- ☒ JSON file
- ☐ XML document

✔ **Correct**

Correct. In a REST API, the client usually receives a JavaScript Object Notation (JSON) file as a response from the web service after sending a request.

4. What term is used to describe a structured collection of data?

1 / 1 point

- ☒ Data Set
- ☐ Data Grid
- ☐ Data Store
- ☐ Data Format

✔ **Correct**

Correct. A data set is a structured collection of data.

5. What type of resources are available for developers on the Data Asset eXchange (DAX) platform?

1 / 1 point

- ☐ Only pre-trained machine learning models
- ☐ Only raw data files without any associated documentation
- ☒ Curated collections of open data sets along with tutorial notebooks
- ☐ Only proprietary data sets with restricted access

✔ **Correct**

Correct. The Data Asset eXchange (DAX) platform provides curated collections of open data sets and tutorial notebooks.

6. What is the primary task of model training in machine learning?

1 / 1 point

- ☒ Identifying patterns in the data
- ☐ Making predictions
- ☐ Preparing the data
- ☐ Evaluating the model

✓ Correct

Correct. The primary task of model training in machine learning is to identify patterns in the data that can be used to make predictions or decisions. During training, the model learns from the input data and adjusts its parameters to minimize the difference between its predictions and the actual outcomes.

7. Which type of machine learning model is primarily used to predict a numeric value?

1 / 1 point

- ☐ Classification
- ☐ Clustering
- ☒ Regression
- ☐ Reinforcement

✓ Correct

Correct. The regression models are used to predict numeric values.

8. Which of the following statements are true regarding deep learning models based on the information provided in the video?

1 / 1 point

- ☒ Deep learning models emulate how the human brain solves problems.
- ☐ Deep learning models are only used for natural language processing tasks.
- ☐ Deep learning models do not require specialized hardware for training.
- ☐ Deep learning models require small data sets for training.

✓ Correct

Correct. Deep learning is a specialized type of machine learning. It refers to a general set of models and techniques that loosely emulate how the human brain solves various problems.

9. Which of the following domains are covered by models available on the Model Asset eXchange?

1 / 1 point

- ☐ Text classification and natural language processing only
- ☒ Object detection, image classification, audio processing, and more
- ☐ Only image classification
- ☐ Only object detection

✓ Correct

Correct. MAX provides models for various domains covering many applications, including object detection, image classification, audio processing, and more.

10. What tool is used to edit front-end languages like HTML, JavaScript, and CSS in the context of exploring machine learning models?

1 / 1 point

- ☐ Red Hat OpenShift
- ☐ IBM Cloud
- ☐ GitHub
- ☒ CodePen

✓ Correct

Correct. CodePen is a tool used to edit front-end languages like HTML, JavaScript, and CSS in the context of exploring machine learning models.

## WEEK 4/MODULE 4

### GRADED QUIZ 1

#### ✔ Congratulations! You passed!

Grade  
received 90%

Latest Submission  
Grade 90%

To pass 70% or  
higher

Go to next item

1. In which formats can you share a notebook with other users?

1 / 1 point

- ☐ ppt
- ☐ xls
- ☒ PDF
- ☐ doc

✔ Correct

Correct! You can export the notebook to a PDF or HTML file that can be shared with anyone.

2. What is a feature of a JupyterLab?

1 / 1 point

- ☒ Interactive control of the notebook cells and output
- ☐ Allows support of a single HTML file format
- ☐ Allows access to a single Notebook
- ☐ Is not an open source

✔ Correct

Correct! JupyterLab allows interactive control of the notebook cells and output.

3. Which option gives the output of the current highlighted code cell in Jupyter Notebook?

1 / 1 point

- ☒ Run -> Run Selected Cells
- ☐ Edit -> Run
- ☐ Run -> Run All Cells
- ☐ Run -> Render All Markdown Cells

✔ Correct

Correct! To get an output from a code cell, place your cursor in the cell, click on Run on the top of your screen, and click **Run Selected Cells** to run the current highlighted cells.

4. How can you switch the kernel in the Jupyter Notebook?

1 / 1 point

- ☐ From the Edit menu on top of the notebook
- ☐ By clicking on the Code option
- ☒ By clicking the top right of the notebook
- ☐ By clicking the top left of the notebook

✔ Correct

Correct! You can also run other kernels by choosing the preferred kernel on the launch page or by clicking the top right and picking the kernel from the drop-down menu.

5. Which of the following can be used to work with Jupyter Notebook?

1 / 1 point

- ☐ Notepad++
- ☐ Explorer
- ☐ Notepad
- ☒ Browser

✔ Correct

Correct! The kernel executes the code and returns the result to the client browser for display.



6. What do you use to convert the ipynb file to other formats?

1 / 1 point

- ☐ Client Browser
- ☐ Preprocessor
- ☐ Postprocessor
- ☒ Exporter

✓ **Correct**  
Correct! The Exporter converts the notebook into the new file format using the NB convert tool.

7. What is one of the features of JupyterLite?

1 / 1 point

- ☐ Default kernel is SQLite
- ☒ Can create interactive graphics and visualizations
- ☐ Requires a dedicated server
- ☐ It is not a static website

✓ **Correct**  
Correct! JupyterLite can be used to create interactive graphics and visualizations, because it supports many visualization libraries like Altair, Plotly, and ipywidgets.

8. What is one of the features of GoogleColab?

0 / 1 point

- ☐ GoogleColab projects are stored on the local machine
- ☒ GoogleColab Jupyter notebooks execute on a server
- ☐ You need to set up and install GoogleColab Jupyter notebooks
- ☐ You can clone GitHub projects and execute them in GoogleColab

✗ **Incorrect**  
Incorrect. Review the Additional Cloud-based Jupyter Environments video.

9. Which environment is used for creating and modifying Jupyter Notebooks on a local device?

1 / 1 point

- ☐ SQLite
- ☐ Python
- ☒ Visual Studio Code
- ☐ GoogleColab

✓ **Correct**  
Correct! JupyterLab and VS Code are popular environments for creating and modifying Jupyter notebooks on a local device.

10. How can you execute Python code in Visual Studio code?

1 / 1 point

- ☒ By installing Python extensions
- ☐ By clicking on Download as
- ☐ By selecting Code from the drop-down menu
- ☐ By selecting Markdown from the drop-down menu

✓ **Correct**  
Correct! To execute Python code in VS Code, you must install Python extensions.

## MODULE 5

### GRADED QUIZ 1

#### ✓ Congratulations! You passed!

Grade  
received 100%

Latest Submission  
Grade 100%

To pass 70% or  
higher

[Go to next item](#)

1. Which of the following is True about R language ?

1 / 1 point

- ☐ R functions require lots of coding compared to other Data science tools
- ☒ R supports importing of data from different sources like flat files, databases
- ☐ R requires installation of additional libraries to handle data analysis
- ☐ R is used for statistical inference and does not support visualization

✓ Correct

Correct! R supports importing of data from different sources like flat files, databases, web, and statistical software such as SPSS and STATA.

2. Which R library is used for machine learning?

1 / 1 point

- ☐ ggplot
- ☐ dplyr
- ☐ stringr
- ☒ caret

✓ Correct

Correct! Caret is used for machine learning.

3. Which function in ggplot adds a title to the plot?

1 / 1 point

- ☒ ggtitle
- ☐ ggplot
- ☐ geom point
- ☐ library

✓ Correct

Correct. The ggtitle function adds the title to the plot.

4. Which function is used to specify appropriate names for both axes in a plot?

1 / 1 point

- ☐ ggplot
- ☐ library
- ☐ geom point
- ☒ xlab and ylab

✓ Correct

Correct. The labs arguments are used to specify appropriate names for both axes in a plot.

5. What is a copy of a repository?

1 / 1 point

- ☐ Working directory
- ☐ SSH protocol
- ☐ Pull request
- ☒ Fork

✓ Correct

Correct! A fork is a copy of a repository.

6. How does GitHub provide streamline testing and delivery?

1 / 1 point

- ☐ With merging
- ☐ With branching
- ☐ With reviewing and including comments
- ☒ With Built-in Continuous Integration (CI) and Continuous Delivery (CD)

✓ Correct

Correct. GitLab provides streamline testing and delivery with Built-in Continuous Integration (CI) and Continuous Delivery (CD).

7. In GitHub, what is an organization?

1 / 1 point

- ☐ A name you must specify for your repository
- ☒ A collection of user accounts that owns repositories
- ☐ A path to the code files in a project
- ☐ The top level of a repository tree

✓ Correct

Correct! An organization is a collection of user accounts that owns repositories.

8. On which tab in the Repository can you see all the source files?

1 / 1 point

- ☒ Code
- ☐ Pull requests
- ☐ Issues
- ☐ Projects

✓ Correct

Correct! All the source files reside in the Code tab.

9. How do you save changes in the Repository?

1 / 1 point

- ☐ Create new file
- ☐ Save changes
- ☐ Add file
- ☒ Commit changes

✓ Correct

Correct! The "commit changes" is used to save your changes to the repository.

10. Which option enables you to add a file to the Repository from your local machine?

1 / 1 point

- ☐ Create new file
- ☒ Upload files
- ☐ Commit changes
- ☐ New Repository

✓ Correct

Correct. You can upload a file from your local system into the repository, click Add File and choose the Upload files option.

## FINAL EXAM

### GRADED QUIZ

**Your grade: 91.66%**

Your latest: 91.66% • Your highest: 91.66% • To pass you need at least 70%. We keep your highest score.

Next item →

1. SQL is what type of query language?

1 / 1 point

- ☐ Contextual
- ☒ Database
- ☐ Cypher
- ☐ Access

✓ Correct

2. Is Keras a machine learning or deep learning library for Python?

1 / 1 point

- ☒ Yes
- ☐ No

✓ Correct

3. What type of model is used to predict a target variable with two classes?

1 / 1 point

- ☐ Clustering model
- ☐ Regression model
- ☒ Classification model

✓ Correct

4. What is Data Asset Management also known as?

1 / 1 point

- ☐ Data strategy
- ☐ Data stewardship
- ☒ Data governance
- ☐ Perimeter security

✓ Correct

5. What open source tool was developed and built by statisticians?

1 / 1 point

- ☐ VSCode
- ☐ Jupyter Notebooks / JupyterLab
- ☒ RStudio
- ☐ Spyder
- ☐ Notepad

✓ Correct

6. How is the R programming language different than Python?

1 / 1 point

- ☐ It's primary objective is deployment and production
- ☒ It was built by statisticians and their specific language
- ☐ It is a general purpose language

✓ Correct

7. Editing and execution of R code is best done in what tool?

0 / 1 point

- ☐ Jupyter Notebooks / JupyterLab
- ☒ RStudio
- ☐ Notepad
- ☐ Both A and B are correct

✗ Incorrect

8. What type of file format is a Jupyter Notebook?

1 / 1 point

- ☒ .ipynb
- ☐ .ppt
- ☐ .xlsx
- ☐ .doc

✓ Correct

9. What is the purpose of ggplot library in R?

1 / 1 point

- ☐ Machine Learning
- ☒ Data Visualization
- ☐ String Manipulation
- ☐ All of the above
- ☐ Data Manipulation

✓ Correct

10. Which feature of GitHub is a way of proposing changes to main branch?

1 / 1 point

- ☐ Fork
- ☐ All of the above
- ☐ Cloning
- ☒ Pull Request

✓ Correct

11. Which among the following is not a feature of Jupyterlite ?

1 / 1 point

- ☐ JupyterLite does not require a dedicated Jupyter server.
- ☐ All of the above
- ☐ Can deploy as a static website
- ☒ Does not execute in browser.
- ☐ It can be used to create interactive graphics and visualizations because it

✓ Correct

12. Which among the following is not a feature of Googlecolab?

1 / 1 point

- ☒ Upload and share with setup and installation
- ☐ Can be stored on Google drive
- ☐ Clone from GitHub and execute in GoogleColab
- ☐ Execute on browser

✓ Correct