Quiz-Course 2. IBM Tools for Data Science (Coursera)

Week 1- Practice Quiz - Data Science Tools

1. Which Data Science category do you extract, transform, and load data?
*   Data Management
*   **Data Integration and Transformation**
*   Model Building
*   Data Visualization

2. Which open-source tool is the <u>standard</u> used for **code asset management**?
*   **Git**
*   Apache PredictionIO
*   ModelDB
*   MySQL

3. Which open source tool supports more than a hundred different programming languages through "kernels"?
*   RStudio
*   Apache Spark
*   Spyder
*   **Jupyter**

4. Which commercial tool can be used to define and execute data integration processes in a **spreadsheet-style**?
*   Oracle Database
*   SPSS Modeler
*   RStudio
*   **Watson Studio Desktop**
    ➔ *Watson Studio Desktop includes a component 'Data Refinery' (allows for the definition/execution of data integration processes in a spreadsheet-style.*

5. Which service offers a **document data structure** format?
*   Data Refinery
*   DataMeer
*   Db2
*   **JSON**

Week 1- Graded Quiz – Data Science Tools

1. Which Data Science category enables you to present data in a form of charts, plots, and maps?
*   Data integration and Transformation
*   Data management
*   **Data visualization**
*   Model building

2. Which Data Science category uses tools like **Fiddler** to **track the performance** of deployed models in a production environment?
*   Data integration and Transformation
*   Data visualization
*   **Model Monitoring**

3. Which open-source tool is used for **model deployment**?
*   Git
*   ModelDB
*   **Apache PredicitonIO**
*   MySQL

4. Which open-source tool is used to **version data and annotate the data with metadata**?
*   Apache AirFlow
*   Hue
*   Kubernetes
*   **Apache Atlas**

5. What feature differentiates **Apache Zappelin** from Jupyter Notebooks?
*   Ability to open different types of files
*   Ability to unity programming, execution and debugging into one tool
*   **Plotting capability**
*   Ability to unify documentation and code in a single document
    ➔ *Apache Zappelin has built-in inline plotting capability, whereas Jupyter Notebooks need libraries such as matplotlib*

6. Which of the following tool is used as a **stream processing image**?
*   **Apache Flink**
*   Apache Spark
*   Zeppelin
*   KNIME

7. Which **Business Intelligence** commercial tool creates live dashboards?
- SPSS Modeler
- Informatica
- Talend
- **Tableau**

8. Which commercial tool will help you **build a machine learning model**?
- InfoSphere DataStage
- **SPSS Modeler**
- IBM Cognos Analytics
- Microsoft DQL Server

9. Which fully integrated visual tool is <u>not deployed by a cloud</u> service provider and needs to be downloaded and installed?
- Watson Studio
- Watson OpenScale
- Microsoft Azure Machine Learning
- **H2O Driverless AI**
  - ➔ *All these four are cloud-based fully integrated visual tool. **H2O Driverless AI**: cloud provider does not do operations and maintenance.*

10. Which of the following tool is part of **IBM Watson Studio**?
- Informatica
- Cloudant
- DataMeer
- **Data Refinery**

Week 2- Practice Quiz – Language of Data Science

1. Which of the following options determines your choice of language to learn in data science?
- Size of organization
- Duration of the problem
- **Problem to be solved and for who are you solving for**
- Number of people who have the problem
  - ➔ *Your choice of language to learn depends on your needs, what the problem to be solved is and on the role of the person with the problem.*

2. In Python, which library is used for Artificial Intelligence?
- NumPy
- Pandas
- **TensorFlow**
- SciPy

3. What are the differences between Python and R languages?
- Python and R both are focused on a set of values
- Python has a different set of licenses than R.
- **Python is open source (OSI), and R is free software (FSF)**.
- Python is free, while R is chargeable.

4. What is the primary purpose of SQL?
- **To query and handle structured data**
- To manage unstructured data
- To process language
- To translate from math to code

5. Which of the following language is an **object-oriented programming language**?
- **Java**
- Scala
- Julia
- C

Week 2- Graded Quiz- Languages of Data Science

1. What does Python use for Natural Language Processing?
- **Natural Language Toolkit (NLTK)**
- TensorFlow
- Keras
- Pandas

Quiz-Course 2. IBM Tools for Data Science (Coursera)

2. In Python, which library <u>is not used</u> for Data Science?
- Pandas
- SciPy
- **Caffe**
- Natural Language Toolkit
  - ➔ *Caffe is a deep learning framework in C++ made with expression, speed, and modularity in mind.*

3. Which of the following is developed using the **R language**?
- Automation
- Relational DBMS
- **Graphing**
- Image Processing

4. Which statement is true for Open source and free software?
- **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
- Free software is more business focused, while free software is more focused on a set of values
- None of the above

5. Which of the following is a **global R community**?
- **useR**
- RealPython
- HackerEarth
- PyLadies

6. Is the following statement true or false: **R integrates well with C and C++**
- **True**
- False
  - ➔ *R integrates well with other computer languages like C and C++, Java, .NET (dot net), and Python*

7. Which of the following <u>is not</u> an SQL language element?
- **Spreadsheets**
- Predicate
- Clauses
- Querie
  - ➔ *SQL language elements include: Clauses, Expressions, Predicates, Queries, and Statements*

8. Why is SQL an American National Standards Institute (or AN-see) standard?
- SQL can considerably speed up workflow executions
- The SQL language is subdivided into several language elements
- **SQL knowledge applies to multiple databases**
- SQL manages data

9. Which **Scala** program is used for <u>data science</u>?
- **Apache Spark**
- Caffe
- Weka
- MangoDB

10. Which is the language used to develop the **Hadoop framework**?
- C++
- Python
- Ruby
- **Java**
  - ➔ *Hadoop is an application of Java which manages data processing and storage for big data applications running in clustered systems.*

Week 3- Practice Quiz – Libraries, APIs, Data Sets, Models

1. Which library offers data structures and tools for effective <u>data cleaning, manipulation, and analysis</u>?
- Matplotlib
- **Pandas**
- Seaborn
- Scikit-learn

2. What is an API?
- Translator
- Data
- **Interface**
- Web service

➔ *API (Application Programming Interface) is the interface used to communicate between software components*

3. What is the best way to represent the **network data**?
- Tabular format
- Comma-separated values
- **As a graph**
- In a tree-like structure
    ➔ *Connections between people on a **social networking website** are often represented as a graph.*

4. What is the primary purpose of the **Data Asset eXchange (DXA)**? Select all that apply.
- To collect datasets with image only
- To organize datasets into tables
- **To collect datasets that are high quality**
- **To keep datasets that have clearly defined license and usage terms**
    ➔ ***DXA** aims to foster data sharing and collaboration by keeping datasets available under a community Data License Agreement (CDLA)*

5. Which of the following are machine learning models? Select all that apply
- **Supervised Learning**
- **Reinforced Learning**
- **Unsupervised Learning**
- Model training

6. Which of the following elements are used to make a model? Select all that apply
- Kubernetes
- Docker
- **Computer resources**
- **Domain expertise**

Week 3- Graded Quiz- – Libraries, APIs, Data Sets, Models
1.Which library is used for machine learning?
- Matplotlib
- **Scikit-learn**
- Numpy
- Pandas

2.Which deep learning library in Python is used for **experimentation**?
- **PyTorch**
- Pandas
- Seaborn
- Numpy
    ➔ *PyTorch is used for experimentation, making it simple for researchers to test ideas (vs. TensorFlow a low level library for large-scale production of DL models*

3.Which API can be used with TensorFlow?
- Seaborn
- Pandas
- C language
- **Julia**
    ➔ *Julia, Matlab, and R volunteer-developed APIs can be used with TensorFlow.*

4. What does T stand for in REST API?
- **Transfer**
- Tensor
- Transport
- Transform
    ➔ *REST API: stands for Representational State Transfer*

5. Which of the following datasets is considered **open data**?
- Pricing data
- Customer data
- Commercially sensitive
- **Government data**

6. Which license stipulates that the **modified version** of the data should be published **under the same license terms** as the original data?
- **CDLA-Sharing**
- CDLA-Permissive
- CDLA-Dataset
- CDLA-License

➔ Both **CDLA-Sharing** license and **CDLA-Permissive** license grant you permission to use and modify the data. **CDLA-Sharing**: if you publish your modified version of the data, you must do so under the same license terms as the original data. **CDLA-Permissive**, you aren't required to share changes to the data

7. Which tab on the IBM developer web page enables you to open the **Data Asset eXchange** page?
- Topics
- Community
- **Open Source at IBM**
- Products & Services

8. Which tab on the **Data Asset eXchange page** enables you to view all the Jupyter Notebooks?
- Jobs
- Manage
- Overview
- **Assets**

9. Which machine learning model is used to solve regression and classification problems?
- Model training
- **Supervised learning**
- Unsupervised learning
- Reinforcement Learning

10. On the **MAX object detector page**, which online tool is used by developers to edit front-end languages?
- **CodePen**
- Node-Red Flow
- Serverless app
- Extract Prediction
   ➔ CodePen: online tool to edit Front-end languages like HTML, JavaScript, and CSS.

Week 4- Practice Quiz – Jupyter Notebooks and JupyterLab

1.What functionality is provides by the Jupyter Notebook file to Data Scientists?
- Runs the code and generates the output in a separate file
- Allows to code in only one language
- **Allows to combine descriptive text, code blocks, and code output in a single file**
- Does not allow to share the file in any format

2.What is the purpose of the kernel in the Jupyter Notebook?
- Exports the code to a different file
- Writes the code
- Translates the code to a different language
- **Executes the code**
   ➔ A **notebook kernel** is a computational engine that executes the code contained in a Notebook file.

3.What are the elements of **Jupyter's architecture**?
- Server and client
- Two kernels and a server
- **Kernel and client**
- Two clients and a server
   ➔ Jupyter implements a two-process model with a kernel and a client

4. What is the default kernel of **JupyterLite**?
- SQLite
- Juila
- **Pyolite**
- No kernel

5. What is the purpose of **Anaconda**?
- **Allows to install new packages without using a command line interface**
- Allows creating interactive visualizations
- Allows execution of code written in any language
- Allows to install packages using a command line interface only

Week 4- Graded Quiz - Jupyter Notebooks and JupyterLab

1.In which formats can you **share a notebook** with other users?
- doc
- xls
- **PDF**
- ppt
   ➔ You can you share a notebook with other users in **PDF or HTML**

2. What is a feature of a **JupyterLab**?
- Is not an open source
- Allows access to a single Notebook
- **Interactive control of the notebook cells and output**
- Allows support of a single HTML file format

3.Which option gives the output of the current highlighted code cell in **Jupyter Notebook**?
- **Run-> Run Selected Cells**
- Edit-> Run
- Run-> Run All Cells
- Run-> Render All Markdown Cells

4.How can you switch the kernel in the **Jupyter Notebook**?
- **By clicking the top right of the notebook**
- From the Edit menu on top of the notebook
- By clicking on the Code option
- By clicking the top left of the notebook

5.Which of the following can be used to work with **Jupyter Notebook**?
- Notepad++
- Notepad
- Explorer
- **Browser**
  - ➔ *The kernel executes the code and return the results to the client browser to display.*

6.What do you use to **convert** the ipynb file to other formats?
- **Exporter**
- Preprocessor
- Client Browser
- Postprocessor
  - ➔ *The Exporter converts the notebook into the new file format using the NB convert tool.*

7.What is one of the features of **JupyterLite**?
- Requires a dedicated server
- **Can create interactive graphics and visualizations**
- It is not a static website
- Default kernel is SQLite
  - ➔ *JupterLite can 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**?
- **You can clone GitHub projects and execute them in GoogleColab**
- You need to set up and install GoogleColab Jupyter notbooks
- GoogleColab projects are stored in the local machine
- GoogleColab Jupyter notebooks execute on a server

9.Which environment is used for creating and modifying Jupyter Notebook on a **local device**?
- GoogleColab
- Python
- **Visual Studio Code (VS Code)**
- SQLite
  - ➔ *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 (VS Code)**?
- By clicking on Download as
- **By installing Python extensions**
- By selecting Markdown from the drop-down menu
- By selecting Code from the drop-down menu

Week 5 -Practice Quiz - RStudio
1.Where can you type R commands in RStudio?
- History
- Files
- Code Editor
- **Console**

Quiz-Course 2. IBM Tools for Data Science (Coursera)

2.Which R library will you use for data visualization such as histogram, barcharts and scatterplot? Select all that apply
- **Plotly**
- Lattice
- Leaflet
- **Ggplot**
  - ➔ *Ggplot: Histograms, bar charts, scatterplots. Plotly for **web-based** data visualization that can be displayed and saved as individual HTML files*

3.Which is the command used to install packages in **R**?
- Library(packages)
- Installpackages()
- **Install.packages()**

Week 5- Practice Quiz -GitHub

1.Which term describes the folders set up for version control
- SSH Protocol
- Pull request
- **Repository**
- Fork

2.Which tab in your Repository enables a review of changes made before being merged into the main branch?
- **Pull requests**
- Code
- Projects
- Issues

3.Which command is used to **clone** an existing repository?
- git status
- **git clone**
- git reset
- git add

Week 5- Graded Quiz- RStudio & GitHub

1. Which of the following is True about R language?
- **R Supports importing of data from different sources like flat files, databases**
- R Functions require lots of coding compared to other Data Science tools
- R is used for statistical inference and does not support visualization
- R requires installation of additional libraries to handle data analysis
  - ➔ *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**?
- dplyr
- stringr
- **caret**
- ggplot

3.Which function in ggplot adds a title to the plot?
- ggplot
- geom point
- **ggtitle**
- library

4. Which function is used to specify appropriate names for both axes in a plot?
- library
- **xlab and ylab**
- geom point
- ggplot

5. What is a **copy of a repository**?
- Working directory
- Pull request
- **Fork**
- SSH protocol

6. How does GitHub provide streamline testing and delivery?
- **With built-in Continuous Integration (CI) and Continuous Delivery(CD)**
- With reviewing and including comments
- With branching
- With merging

7. In GitHub, what is an **organization**?
- **A collection of user accounts that owns repositories**
- With reviewing and including comments
- A name you must specify for your repository
- A path to the code files in a project
- The top level of a repository tree

8.On which tab in the Repository can you **see all the source files**?
- Projects
- **Code**
- Issues
- Pull requests

9. How do you **save changes in a Repository**?
- Save changes
- Create new file
- Add file
- **Commit changes**

10. What option enables you to **add a file to the Repository** from your local machine?
- New repository
- **Upload files**
- Create new file
- Commit changes

Week 7- Practice Quiz – Watson Studio

1. You want your project in **Cloud Pak for Data** to interact with outside tools like Git repository. Which option under Projects will provide you with this capability?
- **Services & integrations**
- Resource usage
- Environments
- Access control
  - ➔ *Services and integrations, use third-party Integrations so your project can interact with external tools. **Project:** shows projects/ jobs you have created.*

2. After you create an account in Watson Studio, which option should you select to **create a project**?
- **Work with Data**
- Projects
- Take a tutorial
- Learn what's new

3. Which option enables part of the Jupyter notebook to not execute?
- Code
- **Markdown**
- Notification
- Job schedule

4.Which option should you select to see the available runtime environments?
- Access control
- **Templates**
- Resource usage
- Services & integrations
  - ➔ Templates: show the available runtime environment. Resource usage: shows the usage of resources. Access control: controls access with user groups

5.Which option should you select to define the access for personal tokens?
- **Select scopes > repo**
- Third-party integrations > Connect integration
- Manage > Service & Integrations
- Select scopes > personal access token

Quiz-Course 2. IBM Tools for Data Science (Coursera)

1. Which of these is a database query language?
- Julia
- Python
- **SQL**

2. Which of these is a machine learning <u>or</u> deep learning library for Python?
- **Scikit-learn**
- NumPy
- Requests
- Pandas

3. What type of model would you use if you wanted to find the **relationship** between a set of variables?
- **Regression model**
- Clustering model
- Classification model

4. Data governance can also be called?
- **Data asset management**
- Perimeter security
- Data stewardship
- Data strategy

5. Which are the two most used open source tools for data science?
- Notepad
- **RStudio**
- **Jupyter Notebooks / JupyterLab**
- Spyder
- VSCode

6. What tool do most R developers use?
- **RStudio**
- Jupyter Notebooks / JupyterLab

7. Storing data in tables is a function that RStudio provides.
- Yes
- **No**

8. Which of the following does Jupyter notebooks support?
- Data visualizations
- Numerous programming languages
- GitHub
- **All of the above**

9. Which one of these libraries is widely used for data manipulation in R?
- stringr
- **dplyr**
- ggplot
- pandas

10. Which of these options is a develops platform delivered as a <u>single application</u>
- GitHub
- Git
- BitBucket
- **Gitlab**

11. Which of the following is a lightweight tool built from Jupyterlab components?
- **Jupyterlite**
- Spider
- VScode
- GoogleColab
- All of the above

12. Which among the following is **NOT** a feature of Googlecolab?
- Execute on browser
- **Upload and share with setup and installation**
- Can be stored on Google drive
- Clone from GitHub and execute in GoogleColab