IBM

Certification

Data & AI

IBM Cloud Professional Certification Program

Study Guide Series

Exam C1000-154: IBM Watson Data Scientist v1

Contents

Purpose of Exam Objectives

When an exam is developed, Subject Matter Experts work together to define the role the certified individual will fill. They define the tasks and knowledge that an individual would need to successfully perform this job role doe the product or solution. This creates the foundation for the objectives and measurement criteria, which form the basis of the certification exam. Question writers then use these objectives to develop exam questions.

It is recommended that you review these objectives and ask yourself the following questions:

- Do you know how to complete the task in the objective?
- Do you know why that task needs to be done?
- Do you know what will happen if you do it incorrectly?

If you are not familiar with a task, go through the objective, perform that task in your own environment and read more information on the task. If there is an objective on a task, there is a high likelihood that you WILL see a question about it on the actual exam. Review the recommended learning designed to prepare you to take the certification exam.

After reviewing the objectives in this guide and completing your own research, take the assessment exam. While the assessment exam does not indicate which specific questions were answered incorrectly, it does indicate overall performance by section. This is a good indicator of preparedness or if further preparation is warranted.

High-level Exam Objectives

Section 1 - Understand the business problem	
1.1	Help business articulate and define business problems
1.2	Identify analytic techniques to address requirements
Section 2 - Collect and explore the data	
2.1	Identify appropriate data sources
2.2	Collect data
2.3	Assess data quality
2.4	Perform exploratory data analysis
2.5	Connect and ingest all data sources
Section 3 - Prepare the data	
3.1	Preprocess and combine data from various data sources
3.2	Clean and validate the data
3.3	Data integration
3.4	Feature selection and engineering
Section 4 - Build the model	
4.1	Select the right model class and toolset
4.2	Split data
4.3	Create models
Section 5 - Evaluate the model	
5.1	Perform hyperparameter tuning
5.2	Compare the performance of different models
Section 6 - Deploy the solution	
6.1	Understand deployment environment considerations
6.2	Create data pipelines to automate model lifecycle
6.3	Deploy models in a production setting
6.4	Validate model performance to business outcomes
Section 7 - Governance and compliance	
7.1	Govern and manage data
7.2	Govern and manage models
Section 8 - Visualization and Storytelling	

8.1	Utilize appropriate visualizations and tools
8.2	Articulate findings to business community
Section 9 - Strategy and Lifecycle	
9.1	Understand and utilize the Data Science/AI Lifecycle
9.2	Collaborate with IT on technical and data architectures
9.3	Illustrate the value of governed data
9.4	Understand and articulate IBM Cloud Pak for Data value propostion

Detailed Exam Objectives

Section 1 - Understand the business problem

1.1. Help business articulate and define business problems (Item writer – scenario based questions)**.**
SUBTASKS:
1.1.1. Understand the CRISP Methodology
REFERENCES:
https://thinkinsights.net/digital/crisp-dm/#CRISP-DM-Methodology
https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=guide-business-understanding

1.1.2. Explain how IBM Garage Methodology works
REFERENCES:
https://www.ibm.com/garage

1.2. Identify analytic techniques to address requirements.
SUBTASKS:
1.2.1. Align on user intents for a solution
1.2.2. Determine upskill requirements
1.2.3. Assess feasibility of solution(s)
1.2.4. Define key metrics

REFERENCES:
https://www.ibm.com/design/thinking/page/courses/AI_Essentials
https://www.ibm.com/design/ai/team-essentials/
https://www.ibm.com/design/thinking/page/toolkit/activity/ai-essentials-intent
https://www.ibm.com/design/thinking/static/team-essentials-for-ai-workbook-8dc9aadb2cc2dc6343cc5e420b522ca2.pdf
https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=guide-business-understanding

Section 2 – Collect and explore the data

2.1. Identify appropriate data sources.

SUBTASKS:
2.1.1. Understand what data sources are available
https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=understanding-collecting-initial-data
    2.1.1.1. Browse data assets using Watson Knowledge Catalog
REFERENCES:
https://www.ibm.com/docs/en/iis/11.7?topic=assets-finding-viewing-asset-in-catalog
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-discovering-assets

2.1.2. Anticipate additional data sources that might be relevant

https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=integrations-external-data-sets
https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=understanding-assessing-situation

2.2. Collect data.

SUBTASKS:
2.2.1. Add data assets from catalog to project (Watson Knowledge Catalog and Cloud Pak for Data)
https://developer.ibm.com/learningpaths/cloud-pak-for-data-learning-path/find-prepare-and-understand-data/
https://dataplatform.cloud.ibm.com/docs/content/wsj/manage-data/add-data-project.html?audience=wdp
https://www.ibm.com/docs/en/iis/11.7?topic=projects-adding-data-project

2.2.2. Collect additional data
    2.2.2.1. Examples of data collection
https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=data-e-retail-example-initial-collection
    2.2.2.2. Use SQL to fetch data from data warehouse
https://www.w3schools.com/sql/
https://www.ibm.com/cloud/learn/data-warehouse
    2.2.2.3. Scrape data from webpages
https://www.crummy.com/software/BeautifulSoup/bs4/doc/

2.2.2.4. Use Python APIs for external data
https://realpython.com/python-api/

2.3. Assess data quality.

SUBTASKS:

2.3.1. Understand what data quality is

https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=understanding-verifying-data-quality
https://crunchingthedata.com/cs01-check-data-quality/

2.3.2. Analyze data quality in WKC and CPD
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-running-quality-analysis
https://towardsdatascience.com/data-quality-dimensions-in-ibm-watson-knowledge-catalog-79cd0aaf0af2
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=assets-using-rules
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-discovering-assets
(INCLUDES SUBTOPICS)
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=results-data-quality-dimensions-violations
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=results-data-quality-score

2.4. Perform exploratory data analysis (EDA).

SUBTASKS:

2.4.1. Determine steps for EDA
https://www.ibm.com/cloud/learn/exploratory-data-analysis
https://learn.ibm.com/mod/page/view.php?id=168485
https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=preparation-data-overview
https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=understanding-exploring-data
https://www.analyticsvidhya.com/blog/2021/05/exploratory-data-analysis-eda-a-step-by-step-guide/

2.4.2. Use pandas in Jupyter notebook for EDA
https://realpython.com/pandas-python-explore-dataset/
https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/
https://www.kaggle.com/code/kashnitsky/topic-1-exploratory-data-analysis-with-pandas/notebook

2.4.3. Profile and visualize data using Watson tools
https://www.ibm.com/cloud/learn/data-visualization
https://dataplatform.cloud.ibm.com/docs/content/wsj/refinery/visualizations.html?audience=wdp
https://developer.ibm.com/learningpaths/cloud-pak-for-data-learning-path/data-visualization-with-data-refinery/

2.5. Connect and ingest all data sources.

SUBTASKS:

2.5.1. Demonstrate knowledge of ETL process
https://www.ibm.com/cloud/learn/etl

2.5.2. Connect to data sources using Cloud Pak for Data
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-connecting-sources
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-supported-sources
https://dataplatform.cloud.ibm.com/docs/content/wsj/manage-data/metadata-import.html?audience=wdp

Section 3 – Prepare the data

3.1. Preprocess and combine data from various data sources.
SUBTASK(S):
3.1.1. Identify potential issues with data
- 3.1.1.1. Is the data representative of the real world
- 3.1.1.2. Completeness
- 3.1.1.3. Consistency

3.1.2. Employ dimensionality reduction techniques for volume reduction
3.1.3. Transform the data based on model requirements

REFERENCES:
https://developer.ibm.com/articles/data-preprocessing-in-detail/

3.2. Clean and validate the data.
SUBTASK(S):
3.2.1. Describe several methods for replacing missing values in data
3.2.2. Describe several methods for detecting outliers in data
3.2.3. Describe class imbalance and ways to avoid it

3.2.4. Deduplicate data

REFERENCES:

https://developer.ibm.com/articles/data-preprocessing-in-detail/
https://www.ibm.com/blogs/systems/tackling-bias-in-ai/
https://www.ibm.com/docs/en/spss-modeler/18.2.2?topic=nodes-data-audit-node
https://research.ibm.com/publications/the-class-imbalance-problem
https://learn.ibm.com/mod/video/view.php?id=168586
https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=preparation-cleaning-data
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-refining
https://www.ibm.com/garage/method/practices/code/data-preparation-ai-data-science/
https://www.ibm.com/garage/method/practices/reason/prepare-data-for-machine-learning/

3.3. Data integration.

SUBTASK(S):

3.3.1. Choose a method of data integration
- 3.3.1.1. Data consolidation
- 3.3.1.2. Data propagation
- 3.3.1.3. Data virtualization

3.3.2. Demonstrate working knowledge of SQL
- 3.3.2.1. Data management
- 3.3.2.2. Data manipulation

3.3.3. Use a variety of tools to merge data from different sources
- 3.3.3.1. SQL Join
- 3.3.3.2. Pandas Merge
- 3.3.3.3. SPSS Merge Node

REFERENCES:

https://developer.ibm.com/articles/data-preprocessing-in-detail/
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-virtualizing
https://www.ibm.com/docs/en/ssw_ibm_i_71/sqlp/rbafy.pdf
https://www.ibm.com/docs/en/ssw_ibm_i_71/sqlp/rbafy.pdf page 76 (86 in pdf)
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=operations-merge-node
https://pandas.pydata.org/docs/reference/api/pandas.merge.html

3.4. Feature selection and engineering.

SUBTASK(S):

3.4.1. Identify and extract key features

3.4.1.1. SPSS feature selection node
3.4.1.2. Python sklearn feature selection
3.4.1.3. Watson NLP APIs
3.4.1.4. Avoid feature leakage

3.4.2. Describe several methods of feature engineering
3.4.2.1. Encoding
3.4.2.2. Embedding
3.4.2.3. Scaling
3.4.2.4. Dimensionality reduction for model optimization

REFERENCES:
https://www.ibm.com/garage/method/practices/reason/prepare-data-for-machine-learning/
https://cloud.ibm.com/apidocs/natural-language-understanding
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=modeling-feature-selection-node
https://scikit-learn.org/stable/modules/feature_selection.html
https://www.kaggle.com/code/alexisbcook/data-leakage/tutorial

Section 4 – Build the model

4.1. Select the right model class and toolset.
SUBTASK(S):
4.1.1. Demonstrate understanding of different types of machine learning and related algorithms
4.1.1.1. Supervised (Regression/Classification)
4.1.1.2. Unsupervised (Clustering)

4.1.2. Differentiate between machine learning and deep learning and describe when to use each
4.1.3. Select a small number of algorithms based on model requirements or use AutoAI
4.1.4. Select a tool based on algorithm requirements and expertise

REFERENCES:
https://www.ibm.com/cloud/architecture/architecture/practices/evaluate-and-select-machine-learning-algorithm/
https://www.ibm.com/cloud/learn/deep-learning
https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=models-autoai
https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=services-watson-machine-learning

4.2. Split data.
SUBTASK(S):
4.2.1. Partition data into train data and test data
REFERENCES:
https://learn.ibm.com/mod/video/view.php?id=165773
4.2.1.1. Create data splits that are reproducible

REFERENCES:
https://www.oreilly.com/library/view/machine-learning-design/9781098115777/ch06.html#problem-id00022
4.2.1.2. Stratified split in case of imbalanced data

REFERENCES:
https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedShuffleSplit.html

4.2.2. Understand the risk of data leakage for model training
REFERENCES:
https://www.kaggle.com/code/alexisbcook/data-leakage/tutorial
https://www.coursera.org/lecture/python-machine-learning/data-leakage-ois3n

4.2.3. Understand and implement cross-validation
REFERENCES:
https://learn.ibm.com/mod/video/view.php?id=166655

4.3. Create models.
SUBTASK(S):
4.3.1. Implement Supervised Learning: Regression
4.3.1.1. Linear regression
4.3.1.2. Ridge/ Lasso regression
4.3.1.3. Logistic regression
4.3.1.4. Random forest regression

REFERENCES:
https://scikit-learn.org/stable/supervised_learning.html

4.3.2. Implement Supervised Learning: Classification
- 4.3.2.1. K-nearest neighbor (KNN)
- 4.3.2.2. Random forest, decision tree
- 4.3.2.3. Support Vector Machines (SVM)
- 4.3.2.4. Naïve Bayes

REFERENCES:
https://scikit-learn.org/stable/supervised_learning.html

4.3.3. Describe several ensemble methods
- 4.3.3.1. Bagging
- 4.3.3.2. Boosting
- 4.3.3.3. Stacking

https://www.ibm.com/cloud/architecture/architecture/practices/evaluate-and-select-machine-learning-algorithm/

4.3.4. Implement Unsupervised Learning: Clustering
- 4.3.4.1. k-means clustering
- 4.3.4.2. Gaussian Mixture Model

REFERENCES:
https://scikit-learn.org/stable/modules/clustering.html

4.3.5. Implement Deep Learning models
- 4.3.5.1. Deep Neural network
- 4.3.5.2 Recurrent neural network (RNN)
- 4.3.5.3 Convolution neural network (CNN)
- 4.3.5.4 Long short-term memory (LSTM)

REFERENCES:
https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/

4.3.6. Watson Studio on Cloud Pak for Data as a Service

REFERENCES:
https://dataplatform.cloud.ibm.com/docs/content/wsj/landings/wsl.html

4.3.6.1 AutoAI

REFERENCES:
https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-overview.html?audience=wdp

4.3.6.2 SPSS Modeler

REFERENCES:
https://dataplatform.cloud.ibm.com/docs/content/wsd/spss-modeler.html?audience=wdp

4.3.6.3 Deep Learning experiment

REFERENCES:
https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml_dlaas.html?audience=wdp

Section 5 – Evaluate the model

5.1. Perform hyperparameter tuning.
SUBTASK(S):
5.1.1. Understand hyperparameters for various algorithms
- 5.1.1.1. Regression
- 5.1.1.2. Classification
- 5.1.1.3. Clustering
- 5.1.1.4. Recommendation engines
- 5.1.1.5. Deep Learning

5.1.2. Describe the trade-offs between underfitting and overfitting a model
Avoid underfitting or overfitting by splitting the data into training, testing, and validation sets

5.1.3. Explain the effect of hyperparameters and hyperparameter tuning
- 5.1.3.1. Tuning is a trial-and-error process
- 5.1.3.2. Tuning is based on the training output loss value
- 5.1.3.3. Learning rate, number of epochs, hidden layers, hidden units, activation
- 5.1.3.4. Functions
- 5.1.3.5. AutoAI hyperparameter optimization

5.1.4. Summarize search algorithms
- 5.1.4.1. Grid Search
- 5.1.4.2. Random Search
- 5.1.4.3. Bayesian Optimization

REFERENCES:
https://www.ibm.com/garage/method/practices/reason/optimize-train-ai-model/
https://www.ibm.com/docs/en/wmla/2.2.0?topic=optimization-hyperparameter-searchalgorithms
https://www.ibm.com/garage/method/practices/reason/evaluate-and-select-machinelearning-algorithm/
https://developer.ibm.com/articles/cc-models-machine-learning/
https://developer.ibm.com/articles/cc-models-machine-learning
https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml_dlaas_hpo.html?linkInPage=true

5.2. Compare the performance of different models.
SUBTASKS:
5.2.1. Different metrics for Regression Models

REFERENCES:
https://scikit-learn.org/0.15/modules/model_evaluation.html#regression-metrics
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=autoai-implementation-details

5.2.2. Different metrics for Classification Models
- 5.2.2.1. Confusion matrix
- 5.2.2.2. AUC measures
- 5.2.2.3. ROC curve
- 5.2.2.4. Precision
- 5.2.2.5. Recall
- 5.2.2.6. F1-score

5.2.3. Choose the best model
- 5.2.3.1. Performance
- 5.2.3.2. Explainability
- 5.2.3.3. Complexity
- 5.2.3.4. Dataset size

REFERENCES:
https://scikit-learn.org/0.15/modules/model_evaluation.html#classification-metrics
https://learn.ibm.com/mod/video/view.php?id=166785
https://learn.ibm.com/mod/video/view.php?id=166786
https://learn.ibm.com/mod/video/view.php?id=169061&forceview=1

https://towardsdatascience.com/considerations-when-choosing-a-machine-learning-model-aa31f52c27f3
https://developer.ibm.com/articles/the-ai-360-toolkit-ai-models-explained/

Section 6 – Deploy the solution

6.1. Understand deployment environment considerations.

REFERENCES:
https://production-gitops.dev/guides/cp4d/platform/overview/

SUBTASKS:
6.1.1. Understands how to use libraries in Python
    6.1.1.1. Understand how to add custom libraries to Cloud Pak for Data before deploying a model

REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=environments-adding-customization

6.1.2. Know which libraries are available in Cloud Pak for Data by default (e.g. Spark)
REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=environments-spark

6.1.3. Understand resources
    6.1.3.1. Different compute and memory resources (e.g. CPU vs. GPU)
REFERENCES:
https://www.weka.io/blog/cpu-vs-gpu/
    6.1.3.2. Computational Memory
REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=environments-notebook#runtime-scope

6.2. Create data pipelines to automate model lifecycle.

SUBTASKS:
6.2.1. Understand the difference between batch processing and streaming
REFERENCES:
https://www.ibm.com/topics/data-pipeline

6.2.2. Know the different data sources available in Cloud Pak for Data

REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-supported-sources

6.2.3. Managing (reading and writing) to different Cloud Pak for Data Services (Watson Studio, WKC, Data Virtualization)

REFERENCES:
WatsonStudio&WKC: https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=data-adding-analytics-project
DataVirtualization: https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=catalogs-data-virtualization-connection
SPSS Modeler: https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=modeler-supported-data-sources-spss

6.2.4. Automate data processing and model deployment with jobs in Watson Studio

REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=projects-jobs

6.3. Deploy models in a production setting.
SUBTASKS:
6.3.1. Deploy models to Watson Machine Learning
REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=deploying-managing-models-functions

6.3.1.1. Deploy in Watson Machine Learning using notebooks

REFERENCES:
ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=notebooks-watson-machine-learning-python-client-example

6.3.1.2. Manage models with Watson Machine Learning
REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=assets-deployment-spaces

6.3.1.2.1. Understand CI/CD
REFERENCES:
https://mlops-guide.github.io/MLOps/CICDML/

6.4. Validate model performance to business outcomes.
SUBTASKS:
6.4.1. Understand application testing methods.
6.4.1.1. A/B Testing.
REFERENCES:
https://www.optimizely.com/optimization-glossary/ab-testing/

6.4.1.2. Multivariate testing.
REFERENCES:
https://www.optimizely.com/optimization-glossary/multivariate-testing/

Section 7 – Governance and compliance

7.1. Govern and manage data.
SUBTASKS:
7.1.1. Understand the governance artifacts in Watson Knowledge Catalog
7.1.1.1. Categories
7.1.1.2. Business Terms
7.1.1.3. Data Classes
7.1.1.4. Reference Data Sets
7.1.1.5. Classifications
7.1.1.6. Policies
7.1.1.7. Governance Rules
7.1.1.8. Data Protection Rules

REFERENCES:
https://docs.openshift.com/container-platform/4.8/authentication/understanding-and-creating-service-accounts.html
https://www.ibm.com/docs/en/cloud-paks/1.0?topic=service-overview

7.1.2. Apply data protection to data
7.1.2.1. Obfuscating vs redacting
7.1.2.2. Access roles and permissions

REFERENCES:
http://www.redbooks.ibm.com/redbooks/pdfs/sg248452.pdf (see 8.4)
https://www.ibm.com/docs/en/cloud-paks/cp-integration/2021.4?topic=management-adding-users-in-cloud-pak-platform-ui

7.2. Govern and manage models.
SUBTASKS:
7.2.1. Manage model deployments
- 7.2.1.1. Manage
- 7.2.1.2. Update
- 7.2.1.3. Scale
- 7.2.1.4. Champion-Challenger model (optional)

REFERENCES:
https://aifs360.mybluemix.net/
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=functions-managing-deployments (INCLUDES SUBTOPICS)
https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=openscale-quality-metrics-overview (INCLUDES SUBTOPICS)
https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml-manage-models.html?audience=wdp
https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/factsheets-model-inventory.html?audience=wdp
https://www.ibm.com/docs/es/sc-and-ds/8.0.0?topic=steps-champion-challenger-overview
https://learning.oreilly.com/library/view/ibm-spss-modeler/9781849685467/ch08s04.html#ch08lvl2sec268

7.2.2. Evaluate model bias
- 7.2.2.1. Bias
- 7.2.2.2. Fairness
- 7.2.2.3. Debiasing
- 7.2.2.4. Explainability

REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=openscale-fairness-metrics-overview (INCLUDES SUBTOPICS)
https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=insights-debiasing-options
https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=insights-explaining-transactions

7.2.3. Evaluate model drift
REFERENCES:
https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=openscale-drift-detection (INCLUDES SUBTOPICS)

REFERENCES: https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=openscale-get-model-insights

Section 8 – Visualization and Storytelling

8.1. Utilize appropriate visualizations and tools.
SUBTASKS:
8.1.1. Implement visualization using tools
    8.1.1.1. Cognos Dashboards
REFERENCES: https://yourlearning.ibm.com/activity/ITS-6X337G

    8.1.1.2. Opensource visualizations
        8.1.1.2.1. matplotlib
        8.1.1.2.2. seaborn
        8.1.1.2.3. plotly
        8.1.1.2.4. ggplot

REFERENCES:
https://matplotlib.org/3.5.2/tutorials/introductory/usage.html
https://seaborn.pydata.org/introduction.html
https://plotly.com/python/getting-started/
https://yhat.github.io/ggpy/docs.html

8.1.2. Employ the type of visualization
    8.1.2.1. Box plots
    8.1.2.2. Bar graphs
    8.1.2.3. Line graphs
    8.1.2.4. Histogram
    8.1.2.5. Scatter plot
    8.1.2.6. Tree map
    8.1.2.7. Heat map
REFERENCES:
https://python-graph-gallery.com/
https://www.ibm.com/analytics/data-visualization
https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15
https://chartio.com/learn/charts/how-to-choose-data-visualization/
https://datavizcatalogue.com/
https://learn.ibm.com/mod/page/view.php?id=168515

8.2. Articulate findings to business community (scenario-based question).

SUBTASKS:

8.2.1. Match data literacy of your audience
- 8.2.1.1. Color choice
- 8.2.1.2. Label usage
- 8.2.1.3. Clutter

REFERENCES:

https://www.tableau.com/learn/articles/data-visualization-tips
https://learning.oreilly.com/videos/engaging-audiences-with/9781491909959/9781491909959-oreillyvideos2092980/
https://www2.insightsoftware.com/dashboard-design-guide/using-the-right-visualizations/

8.2.2. Communicate using stories

REFERENCES:

https://yourlearning.ibm.com/activity/ITS-6X337G?planId=PLAN-8BBC7D459473&sectionId=SECTION-A (Unit 1 and Unit 4)
https://www.lucidchart.com/blog/how-to-tell-a-story-with-data
https://www.stonybrook.edu/commcms/alda-center/thelink/posts/Storytelling%20with%20Data.php
https://www.ibm.com/design/research/research-in-practice/find-the-story/
https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=stories-

Section 9 – Strategy and Lifecycle

9.1. Understand and utilize the Data Science/AI Lifecycle.

SUBTASKS:

9.1.1. Understand the AI Ladder
9.1.2. Explain the challenges adopting AI
9.1.3. Assess progress in infusing AI into the organization
9.1.4. Understand the stages of AI lifecycle (Oreilly-Chpt 4)
9.1.5. Understand design thinking for modern organizations (Oreilly-Chpt 3/IBM)

REFERENCES:

https://www.ibm.com/downloads/cas/O1VADKY2 - Chapter 1
https://ibm-cloud-architecture.github.io/refarch-data-ai-analytics/data/
https://www.oreilly.com/library/view/operationalizing-ai/9781098101329/
https://www.ibm.com/design/thinking/page/framework

9.2. Collaborate with IT on technical and data architectures.

SUBTASKS:

9.2.1. Understand the relationship between data and cloud architectures
9.2.2. Define the relationship between data architecture and AI adoption

REFERENCES:

https://www.ibm.com/downloads/cas/O1VADKY2
Manda, H., Srinivasan, S., Rangarao, D., IBM Cloud Pak for Data, Packt Publishing, Nov. 2021 – Chapter 5 & 6

9.3. Illustrate the value of governed data.

REFERENCES:

https://learn.ibm.com/course/view.php?id=4481
https://dataplatform.cloud.ibm.com/docs/content/wsj/landings/wkc.html
https://dataplatform.cloud.ibm.com/docs/content/wsj/getting-started/videos.html#wkc
(need to select videos from Data Steward section)
https://aifs360.mybluemix.net/governance

SUBTASKS:

9.3.1. Articulate value of common, consistent, and trusted data
- 9.3.1.1. Data quality
- 9.3.1.2. Common sourcing
- 9.3.1.3. Consistent transformation

REFERENCES:

https://www.ibm.com/garage/method/practices/manage/establish-data-governance/
https://www.talend.com/resources/what-is-data-governance/
https://www.ibm.com/analytics/data-governance#:~:text=Data%20governance%20solutions%20and%20tools,for%20storage%20and%20access%20purposes.
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=results-data-quality-score

9.4. Understand and articulate IBM Cloud Pak for Data value propostion.

SUBTASKS:

- 9.4.1. Understand the scalability and flexibility of a modern cloud architecture
- 9.4.2. Understand deployment at scale with trust and transparency
- 9.4.3. Explain self-service analytics
- 9.4.4. IBM Cloud Continuous Delivery

REFERENCES:

https://www.ibm.com/docs/en/cloud-paks/cp-data/2.5.0?topic=overview
https://www.ibm.com/garage/method/practices/deliver/tool_continuous_delivery/

Manda, H., Srinivasan, S., Rangarao, D., IBM Cloud Pak for Data, Packt Publishing, Nov. 2021 – Chapter 2 – 5

Next Steps

1. Take the assessment test for IBM Watson Data Scientist v1
2. If you pass the assessment exam, visit pearsonvue.com/ibm to schedule your testing sessions.
3. If you failed the assessment exam, review how you did by section. Focus attention on the sections where you need improvement. Keep in mind that you can take the assessment exam as many times as you would like ($30 per exam); however, you will still receive the same questions only in a different order.