

Oracle Cloud Infrastructure Data Science Professional: Hands-on Workshop

Student Guide
D1104294GC10



Copyright © 2022, Oracle and/or its affiliates.

Disclaimer

This document contains proprietary information and is protected by copyright and other intellectual property laws. The document may not be modified or altered in any way. Except where your use constitutes "fair use" under copyright law, you may not use, share, download, upload, copy, print, display, perform, reproduce, publish, license, post, transmit, or distribute this document in whole or in part without the express authorization of Oracle.

The information contained in this document is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

Restricted Rights Notice

If this documentation is delivered to the United States Government or anyone using the documentation on behalf of the United States Government, the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs (including any operating system, integrated software, any programs embedded, installed or activated on delivered hardware, and modifications of such programs) and Oracle computer documentation or other Oracle data delivered to or accessed by U.S. Government end users are "commercial computer software" or "commercial computer software documentation" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, reproduction, duplication, release, display, disclosure, modification, preparation of derivative works, and/or adaptation of i) Oracle programs (including any operating system, integrated software, any programs embedded, installed or activated on delivered hardware, and modifications of such programs), ii) Oracle computer documentation and/or iii) other Oracle data, is subject to the rights and limitations specified in the license contained in the applicable contract. The terms governing the U.S. Government's use of Oracle cloud services are defined by the applicable contract for such services. No other rights are granted to the U.S. Government.

Trademark Notice

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Inside are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Epyc, and the AMD logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

Third-Party Content, Products, and Services Disclaimer

This documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.

Table of Contents

- Module 1: Welcome to Data Science** **5**
 - Course Overview 6
- Module 2: Introduction and Configuration** **16**
 - Data Science: Introduction 17
 - ADS SDK Overview 29
 - Tenancy Configuration Basics 40
 - Configure a Tenancy with OCI Resource Manager Oracle Cloud Infrastructure 57
 - Networking for Data Science 63
 - Authenticate to OCI APIs 70
- Module 3: Workspace Design and Setup** **79**
 - Projects 80
 - Notebook Sessions 85
 - How to Work with JupyterLab 92
 - Conda Environments: Overview 97
 - Data Science Conda Environments 105
 - Manage Conda EnvironmentsOracle Cloud Infrastructure 119
 - OCI Vault: Introduction 131
 - Using OCI Vault in OCI Data Science 140
 - Code Repositories (Git) 149
- Module 4: Machine Learning Lifecycle** **161**
 - ML Lifecycle: Overview 162
 - Model Deployment 172
 - Access Data 189

Data Preprocessing	207
Introduction to Feature Types	225
Custom Feature Types	235
Data Visualization	247
Model Training	268
Oracle AutoML: Introduction	274
Hyperparameter Tuning: ADSTuner	295
Model Evaluation	302
Model Explanations: Global Explainer	320
Model Explanations: Local and WhatIf	337
Model Catalog: Overview	347
Model Serialization	367
Module 5: MLOps Practices	376
MLOps Architecture	377
Oracle Cloud Infrastructure Data Science Jobs	386
Create and Mange Jobs for Custom Tasks	403
Jobs Monitoring and Logging	408
Module 6: Related OCI Services	415
Spark Applications, Data Flow, and Data Science	416
Oracle Open Data	431
Oracle Cloud Infrastructure Data Labeling	436
Oracle AI Perceptual Services	447
Oracle AI Decision Services	470