

Course Code: W7S166GS

Course Title: Preparation and Transformation of Data with IBM Data Refinery

Description:

Preparation and transformation of data with IBM Data Refinery is the third course in the learning path for professional data scientists that are working with the IBM Cloud Pak for Data platform. The course aims to familiarize data scientists with the data cleansing, and data shaping capabilities of the Data Refinery tool. Data Refinery saves preparation time by quickly transforming large amounts of raw data into consumable, high-quality information that's ready for analytics.

Learners follow the story of Sara (the data scientist), Muneiza (the data engineer), Liam (the data steward), and Tim (the data quality analyst) working in the Data Analytics department of a large health products company. The company plans a marketing campaign around coupons that are issued to customers and wants to better understand customer behavior. But they first need to access and prepare the relevant data for analytics. The team will mainly use IBM Data Refinery for this task.

Follow along with Sara, Muneiza, Liam, and Tim's story as learners create a suitable data set ready for analytics. Learners verify their acquired knowledge by completing several hands-on lab exercises in a remote classroom environment (provided to each learner during the course introduction).

Objectives:

After completing this course, you should be able to:

- Outline the role of the Data Refinery tool in the ModelOps process.
- Use Data Refinery to profile data.
- Construct various data visualizations.
- Use Data Refinery to analyze and transform data into consumable, high-quality information ready for analytics and machine learning.
- Implement Data Refinery management tasks.

Prerequisites:

The prerequisite skills and knowledge include:

- Basic knowledge of data wrangling and the ModelOps process
- Experience working with tables and databases
- Ability to navigate the Cloud Pak for Data and Watson Studio graphical user interfaces
- Practical experience with routine data management tasks

Duration:

7.2 Hrs

Topics:

- Introduction to Data Refinery
- Connect to your data
- Profile and visualize your data
- Analyze and transform your data
- Manage Data Refinery flows

Audience:

Data Engineers, Data Scientists, Data Quality Analysts, Business Analysts