



# SNOWFLAKE DATA SCIENCE

THREE-DAY COURSE

24J22



UNIVERSITY

DATASHEET

## OVERVIEW

This three-day, role-specific course is intended for participants interested in developing skills and experience using Snowflake AI Data Cloud for data science workloads. The participant will gain exposure to the rich features of Snowflake, diverse machine learning datasets, relevant and popular open source ML frameworks and libraries, and model deployment practices that will provide practical skills applicable to data science jobs. This course consists of lectures, demos, labs, and discussions.

## ACQUIRED SKILLS

- Collect and access data from Snowflake Data Marketplace and other sources.
- Manage and architect data lakes and real-time streams.
- Employ Snowflake-recommended best practices for developing or querying semi-structured and other data types.
- Work with supervised and unsupervised machine learning models using some of the most relevant open source frameworks and libraries.
- Formulate data science and machine learning workflows and data pipelines.
- Manage and deploy machine learning models at scale with APIs.
- Visualize and collaborate on machine learning results.

## WHO SHOULD ATTEND

- Data scientists who build and train machine learning models.
- Data scientists and data analysts who use machine learning models to conduct predictive and prescriptive analytics.

## PREREQUISITES

- Basic knowledge of SQL is required.
- Foundational knowledge of databases.
- Python or some other object-oriented programming language.
- A background in data science, machine learning, or statistical modeling is required.
- Completion of “Snowflake Foundations” one-day course or equivalent Snowflake knowledge.

## DELIVERY FORMAT

Instructor-led Public or Private classes are available.

## TOPICS COVERED

### Overview of Data Science with Snowflake

- Introduction to Data Science Workload
- Connecting to Snowflake

### Snowflake Data Storage

- Supported Object Types
- Supported Data Types
- SQL Support
- The Variant Data Type
- Introduction to Unstructured Data

### Acquire Data

- Accessing External Data
- Loading Data into Snowflake
- Accessing Snowflake Data Worldwide with the Data Cloud
- Snowflake ML Functions
- Cortex LLM

### Prepare Data

- What is Snowpark?
- Sampling Data
- Tidying Tables
- Transforming Data with Snowpark
- Leveraging Unstructured Data
- Table Streams and Tasks

### Perform EDA (Exploratory Data Analysis)

- Tools for EDA
- Univariate Regression in Snowflake
- Estimation Functions

## Perform Feature Engineering

- Feature Engineering
- Pandas on Snowflake
- Feature Engineering with Snowpark

## Train Models

- Overview of Machine Learning
- Snowpark ML
- Snowflake Model Registry
- Training Models with Snowpark Stored Procedures
- Auto ML

## Deploy Models

- Batch Scoring
- Python Worksheets
- UDFs
- Stored Procedures
- Snowpark UDFs for Model Inference
- External Functions

## Beyond Deployment: ML Ops

- Improving Runtime Performance
- Monitoring
- Vectorized UDFs
- Monitoring
- ML Ops