

**OA018G: Data science without a Ph.D. Using IBM SPSS Modeler v18.1.1**

Course Code: OA018G

Duration: 1 days

Instructor-led Training (ILT) | Virtual Instructor-led Training (VILT)

**OVERVIEW**

This course focuses on reviewing concepts of data science, where participants will learn the stages of a data science project. Topics include using automated tools to prepare data for analysis, build models, evaluate models, and deploy models. To learn about these data science concepts and topics, participants will use IBM SPSS Modeler as a tool.

**SKILLS COVERED**

- Introduction to data science and IBM SPSS Modeler
- Setting measurement levels
- Exploring the data
- Using automated data preparation
- Partitioning the data
- Selecting predictors
- Using automated modeling
- Evaluating models
- Deploying models

**WHO SHOULD ATTEND?**

- Business Analysts
- Data Scientists
- Participants who want to get started with data science

**PREREQUISITES**

It is recommended that you have an understanding of your business data

**MODULES****Module 1: Introduction to data science and IBM SPSS Modeler**

- Explain the stages in a data-science project, using the CRISP-DM methodology
- Create IBM SPSS Modeler streams
- Build and apply a machine learning model

**Module 2: Setting measurement levels**

- Explain the concept of "field measurement level"
- Explain the consequences of incorrect measurement levels
- Modify a field's measurement level

**Module 3: Exploring the data**

- Audit the data
- Check for invalid values
- Take action for invalid values
- Impute missing values
- Replace outliers and extremes

**Module 4: Using automated data preparation**

- Automatically exclude low quality fields
- Automatically replace missing values
- Automatically replace outliers and extremes

**Module 5: Partitioning the data**

- Explain the rationale for partitioning the data
- Partition the data into a training set and testing set

**Module 6: Selecting predictors**

- Automatically select important predictors (features) to predict a target
- Explain the limitations of automatically selecting features

**Module 7: Using automated modelling**

- Find the best model for categorical targets
- Find the best model for continuous targets
- Explain what an ensemble model is

**Module 8: Evaluating models**

- Evaluate models for categorical targets
- Evaluate models for continuous targets

**Module 9: Deploying models**

- List two ways to deploy models
- Export scored data

**END OF PAGE**