Advanced Data Preparation Using IBM SPSS Modeler (v18.1.1) SPVC

Code: 0E058G-SPVC
URL: View Online

This IBM Self-Paced Virtual Class (SPVC) includes:
- PDF course guide available to attendee during and after course
- Lab environment where students can work through demonstrations and exercises at their own pace

Contains PDF course guide, as well as a lab environment where students can work through demonstrations and exercises at their own pace.

This course covers advanced topics to aid in the preparation of data for a successful data science project. You will learn how to use functions, deal with missing values, use advanced field operations, handle sequence data, apply advanced sampling methods, and improve efficiency.

If you are enrolling in a Self Paced Virtual Classroom or Web Based Training course, before you enroll, please review the Self-Paced Virtual Classes and Web-Based Training Classes on our Terms and Conditions page, as well as the system requirements, to ensure that your system meets the minimum requirements for this course.

Skills Gained

Please refer to course overview

Who Can Benefit

This advanced course is intended for anyone who wants to become familiar with the full range of techniques available in IBM SPSS Modeler for data preparation.

Prerequisites

- Experience using IBM SPSS Modeler including familiarity with the Modeler environment, creating streams, reading data files, exploring data, setting the unit of analysis, combining datasets, deriving and reclassifying fields, and basic knowledge of modeling.
- Prior completion of the *Introduction to IBM SPSS Modeler and Data Science* course is recommended.

Course Details

Course Outline

1: Using functions to cleanse and enrich data
- Use date functions
- Use conversion functions
- Use string functions
- Use statistical functions
- Use missing value functions

2: Using additional field transformations
- Replace values with the Filler node
- Recode continuous fields with the Binning node
- Change a field- s distribution with the Transform node

3: Working with sequence data
- Use sequence functions
- Count an event across records
- Expand a continuous field into a series of continuous fields with the Restructure node
- Use geospatial and time data with the Space-Time-Boxes node

4: Sampling, partitioning and balancing data
- Draw simple and complex samples with the Sample node
- Create a training set and testing set with the Partition node
- Reduce or boost the number of records with the Balance node

5: Improving efficiency
- Use database scalability by SQL pushback
- Process outliers and missing values with the Data Audit node
- Use the Set Globals node
- Use parameters
- Use looping and conditional execution
