Fuzzy Matching Programming Techniques Using SAS® Software

This 1-hour course teaches fuzzy matching techniques using the SOUNDEX (phonetic matching) algorithm, and the SPEDIS, COMPLEV, and COMPGED functions

Course Description

Data comes in all forms, shapes, sizes and complexities. Stored in files and data sets, SAS® users across industries know all too well that data can be, and often is, problematic and plagued with a variety of issues. When a unique and reliable identifier, referred to as a key, is available, users are generally able to search or match records from two or more data sets using programming techniques without problem. But, when a unique identifier is not reliable or does not exist, such as with files containing subscriber names, mailing addresses, and/or misspelled email addresses, where one or more characters are transposed, or are partially and/or incorrectly recorded the problem becomes more complicated. This course will introduce what fuzzy matching is, along with the application of the SOUNDEX (for phonetic matching) algorithm, the SPEDIS, COMPLEV, and COMPGED functions to resolve key identifier issues and to successfully match, merge, and join less than perfect or messy data.

Learning Objectives

- Introduce and describe the search, merge and join process
- Introduce the assortment of possible data issues
- Describe the fuzzy matching process
- Explore the Soundex (for phonetic matching) algorithm
- Explore the SPEDIS function
- Explore the COMPLEV function
- Explore the COMPGED function

Intended Audience

- Students
- Data Analysts and Data Managers
- Programmers and Application Developers
- Statisticians and Statistical Programmers

Length and Format

50-minute Instructor-led with Examples

SAS® Consultant, Application Developer, Programmer, Data Analyst, Educator and Author

Kirk Paul Lafler

Kirk Paul Lafler is an entrepreneur, consultant, programmer, and SAS software user since 1979. Kirk currently works as a consultant, application developer, programmer, data analyst, educator and author; a lecturer and adjunct professor at San Diego State University; an advisor and adjunct professor at the University of California San Diego Extension; and an educator of dozens of SAS, SQL, R and Python courses, seminars, workshops, and webinars to thousands of users around the world.

As the author of several books including PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press. 2019) along with hundreds of papers and articles on a variety of SAS topics; Kirk has served as an Invited speaker, educator, keynote and section leader at SAS conferences and meetings worldwide; and is the recipient of 25 “Best” contributed paper, hands-on workshop (HOW), and poster awards.

Kirk Paul Lafler ♦ KirkLafler@cs.com ♦ https://www.linkedin.com/in/KirkPaulLafler/ ♦ https://www.linkedin.com/in/Order-of-Magnitude-Analytics/