Data Analytics: Concepts, Challenges, and Solutions Using SAS® Software

By Kirk Paul Lafler, sasNerd

Course Description

Data is ubiquitous. With 2.5 quintillion bytes (1 with 18 zeros) of new data being created each day, data literacy – the ability to analyze and interpret data – is an increasingly valuable skill, particularly in the age of big data. Organizations across industries are embracing data analytics resulting in a growing demand for qualified and experienced talent with essential data and analysis skills. Data analytics is the process and practice of analyzing data to identify trends, extract insights, answer questions, and help understand things that were learned that could negate or call into question the assumptions going into the analysis. This course introduces concepts, challenges, and solutions associated with the data ecosystem and data analytics lifecycle's steps: data collection / extraction, cleaning / transformation, data analysis, visualization, and interpretation using SAS® software.

Course Topics

Course topics include:

- Data types:
  - ✓ Big data: Volume, velocity, and variety.
  - ✓ Metadata.
  - ✓ Quantitative (structured) versus qualitative (unstructured) data.
- Data measurement scale: nominal, ordinal, interval, and ratio.
- Data file types: text, CSV, XLSX, SAS7BDAT, and JSON.
- Data collection / extraction techniques.
- Data cleaning techniques:
  - ✓ Identify and remove errors, duplicates, and outliers.
  - ✓ Identify and remove unwanted data points.
  - ✓ Identify and provide structure to data points.
  - ✓ Identify and impute missing data points.
- Data transformation techniques: subset, sort, concatenation, merge, join, and transpose.
- Transform raw data into meaningful, actionable insights:
  - ✓ Descriptive analytics – what happened in the past.
  - ✓ Diagnostic analytics – why it happened.
  - ✓ Predictive analytics – what is likely to happen in the future.
  - ✓ Prescriptive analytics – what the best course of action (or recommendations) might be.
- Data analysis techniques:
  - ✓ PROC CORR – assesses the strength and direction of a linear relationship between pairs of continuous numeric variables.
  - ✓ PROC FREQ – determine if two categorical variables are independent of each other using Chi-square test of independence.
  - ✓ PROC TTEST with CLASS and VAR statements – using independent sample t tests to determine if the means of two independent groups are significantly different.
  - ✓ PROC TTEST with a PAIRED statement – using paired t tests to determine if the means of two paired measurements are significantly different.
  - ✓ Regression analysis – estimate the relationship between a set of variables.
  - ✓ Factor analysis – reduce large number of variables to a smaller number of factors.
  - ✓ Cohort analysis – produces related groups for analysis purposes.

- ✓ Cluster analysis – gain insight into how data is distributed in a dataset.
- ✓ Time series analysis – identify trends and cycles over time.
- ✓ Sentiment analysis – interpret and classify the emotions of qualitative (textual) data.
- ✓ Monte Carlo analysis – produce models of possible outcomes and their probability distributions.
- ▪ Visualization techniques: statistical graphics – histograms, bar charts, line charts, box plots, scatter plots, pie charts, area charts, geographical maps, network graphs, and Venn diagrams.
- ▪ Interpretation techniques: storytelling.

Intended Audience: Programmers, Data Analysts, Data Scientists, Statisticians, and Others wanting to Learn Data Analytics using SAS® software.
Prerequisites: No previous SAS programming experience required.
Delivery Method: Instructor-led workshop with code examples.
Length: 4 Hours
Course Material: e-Course Notes and SAS code are provided to Attendees.

Bio for Kirk Paul Lafler

Kirk Paul Lafler is a SAS, SQL, RDBMS, Excel, and Python educator, developer, programmer, consultant, and data analyst; and works as an adjunct professor at San Diego State University; an advisor and adjunct professor at the University of California San Diego Extension; and an educator teaching SAS, SQL, Python, Excel, and cloud-based technology courses, workshops, and webinars to users virtually around the world. Kirk has nearly 50 years of data and programming experience, and specializes in SAS software, SQL, RDBMS technologies (Oracle, SQL-Server, Teradata, DB2), Python, and other languages and productivity tools. As an author, Kirk's most recent books include Exploratory Data Analysis (EDA) By Example (PB&J Press. Summer 2023), Python ETL By Example (PB&J Press, Fall 2023), PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press. 2019), and other books along with numerous papers and articles on a variety of SAS, SQL, and Python topics. Kirk is actively involved with international, regional, and local SAS, SQL, and Python user groups and conferences as an Invited speaker, educator, keynote, and section leader; and is the recipient of 27 "Best" contributed paper, hands-on workshop (HOW), and poster awards.

Kirk Paul Lafler
sasNerd
SAS / SQL / Python Educator, Developer, Programmer, Consultant, Data Analyst, and Author
KirkLafler@cs.com
@sasNerd
https://www.linkedin.com/in/KirkPaulLafler