Data Analysis Boot Camp

DATA200; 3 Days, Instructor-led

Course Description

Today's organizations face both a promise and a dilemma. The growth in availability and quantity of data, as well as the tools to leverage it, is well understood. Every day buzzwords like "big data," "insights" and "analytics" permeate the pages of our business journals. However, much less available are the actual skills to truly understand and realize the benefits of this explosive growth. The potential is very real, but comprehensive skills can be scarce, and outside consultants are expensive.

Fortunately, you don't need a PhD in data science to achieve the rewards of good data analysis and management. If you have basic familiarity with a tool like Excel, this three-day course can teach you the comprehensive skills and tools to maximize and leverage your data assets.

Course Objectives

- Understanding Data
- Looking at Data
- Modeling Data
- Mining Data
- Using Data

Audience Profile

Anyone involved in operations, project management, business analysis, or management would benefit from this class. This training course is invaluable data analysis training for:

- Business Analyst, Business Systems Analyst, CBAP, CCBA
- Systems, Operations Research, Marketing, and other Analysts
- Project Manager, Program Manager, Team Leader, PMP, CAPM
- Data Modelers and Administrators, DBAs
- IT Manager, Director, VP
- Finance Manager, Director, VP
- Operations Supervisor, Manager, Director, VP
- Risk Managers, Operations Risk Professionals
• Process Improvement, Audit, Internal Consultants and Staff
• Executives exploring cost reduction and process improvement options
• Job seekers and those who want to show dedication to process improvement
• Senior staff who make or recommend decisions to executives

Course Outline

Module 1: Data Fundamentals

• Course Overview and Level Set
• Objectives of the class
• Expectations for the class
• Understanding "real-world" data
• Unstructured vs. structured
• Relationships
• Outliers
• Data growth
• Types of Data
• Flavors of data
• Sources of data
• Internal vs. external data
• Time scope of data (lagging, current, leading)

Lab: Hands-on – Profiling Data

• Data-related Risk
• Common identified risks
• Effect of process on results
• Effect of usage on results
• Opportunity costs, Tool investment
• Mitigators of risks
• Data Quality
• Cleansing
• Duplicates
• SSOT
• Field standardization
• Identifying sparsely populated fields
• How to fix common issues

Lab: Hands-on – Dealing with Duplicates

• Relationships
• Finding common attributes
• 1:N, N:N, 1:1
Lab: Hands-On – Data Relationships using PowerPivot

Module 2: Analysis Foundations

- Statistical Practices: Overview
- Comparing programs and tools
- Words in English vs. data
- Concepts specific to data analysis
- Domains of data analysis
- Descriptive statistics
- Inferential statistics
- Analytical mindset
- Describing and solving problems

Module 3: Analyzing Data

- Statistical Practices: Overview
- Averages in data
  - Mean
  - Median
  - Mode
  - Range
  - Central Tendency
  - Variance
  - Standard deviation
  - Sigma values
  - Percentiles

Lab: Hands-On – Central Tendency LAB: Hands-On – Linear Regression

- Distributions
  - Probability distribution
  - Cumulative distribution
  - Bimodal distributions
  - Skewness of data
  - Pareto distribution
  - Correlation

Lab: Hands-On – Distributions in Consumer Finance Data

- Analytical Graphics for Data
  - Categorical – bar charts
  - Continuous – histograms
  - Time series – line charts
• Bivariate data – scatter plots
• Distribution – box plot

Module 4: Analytics & Modeling

• ROI & Financial Decisions

Lab: Hands-On – Helpful financial metrics in Excel

• Using Financial Data
• Earned Value
• Actual Cost, BAC and EAC
• Expected Monetary Value
• Cost Performance Index
• Schedule Performance Index
• Random Numbers
• Sampling
• Simulation
• Monte Carlo analysis
• Pseudo-random sequences

Demo / Lab – Monte Carlo Analysis in Excel

• Predictive Analytics
• Patterns
• Regression and time series models
• Machine learning
• Tools for predictive analytics

Demo / Lab – Using R for powerful analysis

• Clustering
• Segmentation
• Common algorithms
• K-MEANS
• PAM
• Data Modeling
• Architecture and analysis
• Stages of a data model
• Data warehousing
• Top-down vs. Bottom-up
• Data Warehousing
• Context tables
• Facts
• Dimensions
• Star Schema
• Snowflake Schema

Module 5: Visualizing & Presenting Data

• Goals of Visualization
• Communication and Narrative
• Decision enablement
• Critical characteristics
• Visualization Essentials
• Users and stakeholders
• Stakeholder cheat sheet
• Common missteps

Demo / Lab – Improving a Difficult Report

• Communicating Data-Driven Knowledge
• Alerting and trending
• To self-serve or not
• Formats & presentation tools
• Design considerations