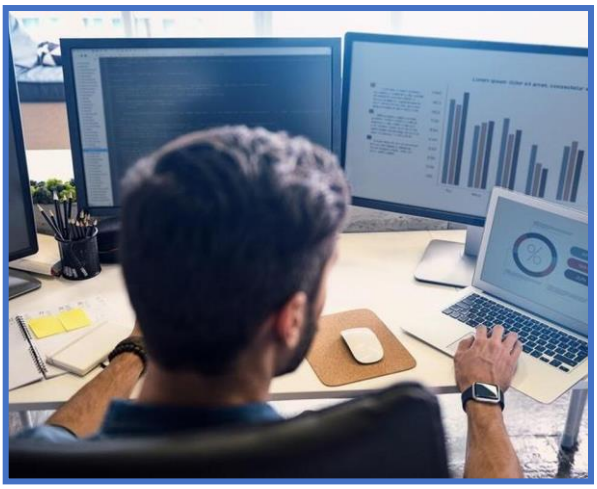
COURSE OVERVIEW GE0749
Data Engineer
(E-Learning Module)



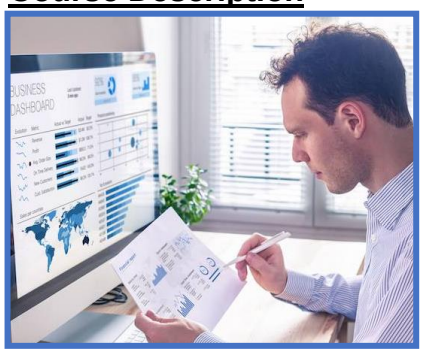
Course Title
Data Engineer (E-Learning Module)

Course Reference
GE0749

Course Format & Compatibility
SCORM 1.2. Compatible with IE11, MS-Edge, Google Chrome, Windows, Linux, Unix, Android, IOS, iPadOS, macOS, iPhone, iPad & HarmonyOS (Huawei)

Course Duration
30 online contact hours
(3.0 CEUs/30 PDHs)

Course Description



This E-learning course is designed to provide participants with a detailed and up-to-date overview of data engineering. It covers the data analysis techniques for engineers, technologists and managers; the basic statistics, data collection and sampling; the sources of data, published data, observational and experimental studies; the general use of graphical and numerical methods; the rules of thumb for drawing histograms; the notation and measures of centrality; the construction of histogram, bar chart and pie chart; the sample and population means; the calculation of the mean, median and mode; the advantages and disadvantages of mean and median; the use of median for asymmetric distributions, the normal distribution and data; the left and right skewed distribution; and the relation between mean and median and measures of variability.

During this course, participants will learn the calculation of the coefficient of variation and the range; the mean and median for grouped data; the box plots, relative frequencies for qualitative data and relative frequencies; the distributions of random variables, basics of probability, frequency interpretation and need of probability; the sum of the probabilities of the sample points; the disjoint or mutually exclusive outcomes, probabilities when events are not disjoint and useful results; the capturing of the population parameter, changing of the confidence level and interpretation of confidence intervals; the hypothesis tests based on a difference means; and the examination of the standard error formula.

Course Objectives

Upon the successful completion of this course, participants will be able to:-

- Apply and gain a comprehensive knowledge on data engineering
- Use data analysis techniques for engineers, technologists and managers
- Explain basic statistics, data collection and sampling, sources of data, published data, observational and experimental studies
- Discuss the general use of graphical and numerical methods as well as the rules of thumb for drawing histograms
- Construct histogram, bar chart and pie chart
- Review notation and measures of centrality including the sample and population means
- Calculate the mean, median and mode as well as discuss the advantages and disadvantages of the mean and the median
- Use median for asymmetric distributions and describe the normal distribution and data
- Review left and right skewed distribution including the relation between mean and median and measures of variability
- Calculate the coefficient of variation and the range, mean for grouped data and median for grouped data
- Describe box plots, relative frequencies for qualitative data and relative frequencies
- Discuss the distributions of random variables, basics of probability, frequency interpretation and need of probability
- Determine the sum of the probabilities of the sample points
- Identify disjoint or mutually exclusive outcomes, probabilities when events are not disjoint and useful results
- Capture the population parameter, change the confidence level and interpret confidence intervals
- Employ hypothesis tests based on a difference means and examine the standard error formula

Who Should Attend

This course provides an advanced overview of all significant aspects and considerations of data engineering for engineers, scientists and managers in the petroleum industry whose job includes a major amount of data analysis. It is especially useful for persons who are responsible for planning and carrying out experimental work or data taking. Further, the course is ideal for laboratory managers, chemists and analysts.

Course Certificate(s)

Internationally recognized certificates will be issued to all participants of the course.

Certificate Accreditations

Certificates are accredited by the following international accreditation organizations: -

- USA International Association for Continuing Education and Training (IACET)

  Haward Technology is an Authorized Training Provider by the International Association for Continuing Education and Training (IACET), 2201 Cooperative Way, Suite 600, Herndon, VA 20171, USA. In obtaining this authority, Haward Technology has demonstrated that it complies with the **ANSI/IACET 1-2013 Standard** which is widely recognized as the standard of good practice internationally. As a result of our Authorized Provider membership status, Haward Technology is authorized to offer IACET CEUs for its programs that qualify under the **ANSI/IACET 1-2013 Standard**.

  Haward Technology's courses meet the professional certification and continuing education requirements for participants seeking **Continuing Education Units** (CEUs) in accordance with the rules & regulations of the International Association for Continuing Education & Training (IACET). IACET is an international authority that evaluates programs according to strict, research-based criteria and guidelines. The CEU is an internationally accepted uniform unit of measurement in qualified courses of continuing education.

  Haward Technology Middle East will award **3.0 CEUs** (Continuing Education Units) or **30 PDHs** (Professional Development Hours) for participants who completed the total tuition hours of this program. One CEU is equivalent to ten Professional Development Hours (PDHs) or ten contact hours of the participation in and completion of Haward Technology programs. A permanent record of a participant's involvement and awarding of CEU will be maintained by Haward Technology. Haward Technology will provide a copy of the participant's CEU and PDH Transcript of Records upon request.

- British Accreditation Council (BAC)

  Haward Technology is accredited by the **British Accreditation Council** for **Independent Further and Higher Education** as an **International Centre**. BAC is the British accrediting body responsible for setting standards within independent further and higher education sector in the UK and overseas. As a BAC-accredited international centre, Haward Technology meets all of the international higher education criteria and standards set by BAC.

Training Methodology

This Trainee-centered course includes the following training methodologies:-

- Talking presentation Slides (ppt with audio)
- Simulation & Animation
- Exercises
- Videos
- Case Studies
- Gamification (learning through games)
- Quizzes, Pre-test & Post-test

Every section/module of the course ends up with a Quiz which must be passed by the trainee in order to move to the next section/module. A Post-test at the end of the course must be passed in order to get the online accredited certificate.

Course Fee

As per proposal

Course Contents

- Data Analysis Techniques for Engineers, Technologists & Managers
- Basic statistics: A Survival Guide
- Basics of Statistics
- Data Collection and Sampling
- Introduction
- Sources of Data
- Published Data
- Observational and Experimental Studies
- Surveys
- Sampling
- Sampling Plans
- Case Study #1
- Quiz #1
- Describing Data
- General Use of Graphical and Numerical Methods
- Graphical Methods
- Histograms
- Rules of Thumb for Drawing Histograms
- Example 2.1
- Construction of a Histogram

- Bar Charts
- Pie Charts
- Example 2.2
- Construction of a Bar Chart and a Pie Chart
- Numerical Methods
- Notation and Measures of Centrality
- The Sample and Population Means
- The Sample Median
- The Sample Mode
- Example 2.3
- Calculation of the Mean, Median and Mode
- The Weighted Mean
- The Mode
- Advantages and Disadvantages of the Mean and the Median
- Use of Median for Asymmetric distributions
- The Normal Distribution
- Describing Data
- A left Skewed Distribution
- A right Skewed Distribution
- Skewness and the Relation between Mean and Median
- Measures of Variability (Dispersion)
- The Sample Variance
- Standard Deviation
- Coefficient of Variation
- Example 2.4
- Calculation of Sample Variance
- Calculation of the Coefficient of Variation and the Range
- Grouped Data
- The Mean of Grouped Data
- The Sample Variance of Grouped Data
- Calculation of the Mean for Grouped Data
- Calculation of the Median for Grouped Data
- Percentiles and Box Plots
- The Interquartile Range
- Boxplot (Box and Whisker plot)
- Box Plot

- Description and Comparison of Box Plots
- Describing Relative Frequencies for Qualitative Data
- Describing Relative Frequencies
- Case Study #2
- Quiz #2
- Distributions of Random Variables
- The Basics of Probability
- Probability
- Frequency Interpretation of Probability
- The Need of Probability
- Sum of the Probabilities of the Sample Points
- Operations on Events
- Disjoint or Mutually Exclusive Outcomes
- Operations on Events
- Probabilities When Events are not Disjoint
- Useful Results
- Probability Distribution
- Bernoulli Distribution
- The Binomial Distribution
- Formula for the Binomial Distribution
- Probability Distribution
- Expectation
- The Normal Distribution
- The “Coin Tossing” Distribution
- Random Sampling
- The “Coin Tossing” Distribution of 10-Toss Data Points
- The “Coin Tossing” Distribution of 25-Toss Data points
- The “Coin Tossing” Distribution of 50-Toss Data points
- The “Coin Tossing” Results for 50,000 data points of 32 Tosses Each
- You will See this Curve much more!
- Normal Distribution
- Normal Distribution Model
- Standardizing with Z scores
- Normal Probability Table
- Normal Probability Examples
- 68-95-99.7 rule

- Case Study #3
- Quiz #3
- Foundations for Inference
- Introduction
- Variability in Estimates
- Point Estimates
- Point Estimates are not Exact
- Standard Error of the Mean
- Point Estimates are not Exact
- Standard Error of the Mean
- Basic Properties of Point Estimates
- Confidence Intervals
- Capturing the Population Parameter
- An Approximate 95% Confidence Interval
- A Sampling Distribution for the Mean
- Changing the Confidence Level
- Interpreting Confidence Intervals
- Nearly Normal Population with known SD
- Case Study #4
- Quiz #4
- Foundations for Inference Hypothesis Testing
- Hypothesis Testing
- Hypothesis Testing Framework
- Decision Errors
- Formal Testing Using p-values
- Two-Sided Hypothesis Testing with p-values.
- Choosing a Significance Level
- Examining the Central Limit Theorem
- Inference for Other Estimators
- Confidence Intervals for Nearly Normal Point Estimates
- Hypothesis Testing for Nearly Normal Point Estimates
- Non-Normal Point Estimates
- When to Retreat
- Sample Size and Power (Special Topic)
- Finding a Sample Size for a Certain Margin of Error
- Power and the Type 2 Error Rate

- Statistical Significance Versus Practical Significance
- Case Study #5
- Quiz #5
- Inference for Numerical Data
- Introduction
- Paired Data
- Paired Observations and Samples
- Inference for Paired Data
- Difference of Two Means
- Point Estimates and Standard Errors for Differences of Means
- Confidence Interval for the Difference
- Hypothesis Tests Based on a Difference in Means
- Summary for Inference of the Difference of Two Means
- Examining the Standard Error Formula
- One-Sample Means with the t distribution
- The Normality Condition
- Introducing the t distribution
- The t distribution as a solution to the Standard Error Problem
- One sample t confidence Intervals
- One Sample t tests
- The t distribution for the difference of two means
- Sampling distributions for the difference in two means
- Two sample t test
- Two sample t confidence interval
- Case Study #6
- Quiz #6
- Comparing many means with ANOVA
- Pooled Standard Deviation Estimate
- Comparing many means with ANOVA
- Is batting Performance Related to Player Position in MLB?
- Analysis of variance (ANOVA) and the F test
- Reading an ANOVA table from software
- Graphical Diagnostics for an ANOVA analysis
- Multiple Comparisons and Controlling Type 1 Error rate
- Analysis of Variance (ANOVA)
- Case Study #7

- Quiz #7
- Introduction to Linear Regression
- Introduction
- Line Fitting, Residuals, and Correlation Beginning with Straight Lines
- Beginning with Straight Lines
- Fitting a line by eye
- Residuals
- Describing Linear Relationships with Correlation
- Fitting a line by Least Squares Regression
- Linear Regression and Calibration Curves
- Applying Linear Regression for Standardizing Methods
- Case Study #8
- Quiz #8
- Revision Example
- Data Collection, Analysis, Interpretation and Communication
- Analysis of Experimental Data
- Standard Error Estimate – for a Single Measurement
- Estimating and Reporting Uncertainty – Repeated Measurements
- Example: Standard Deviation
- Calculation Error Estimates – Propagation of Independent Errors
- Example: Error Propagation
- Linear Regression and the Correlation Coefficient
- Example: Least-squares Linear Regression
- Interpreting Linear, Power, and Exponential Equations
- Example Linear Manipulation of Non-linear Models
- Student's t-test
- Example: One-sided t-test
- Example: Two-sided t-test
- One-way analysis of variance (ANOVA)
- Example: One-way ANOVA
- Case Study #9
- Quiz #9
- Introduction to Statistical Process Control Techniques
- Quality Control Today
- New Demands on Systems Require Action
- So what is Statistical Process Control?

- Where did this Idea Originate?
- What exactly are Process Control Charts?
- What is the Purpose of Control Charts?
- How do they work?
- What's this Relationship Between Variation and Assignable Causes?
- So how are these Normal-Predictable Variance Levels Determined?
- What about these Rules Violations for Determining if a Series of Points within the Control Limits is Unnatural?
- In Control? Out of Control? What's the Point?
- What's this about Capabilities?
- Can any Type of Process Data be Judged Using Control Charts?
- What Specifically do they look like, what are their Key Features and how are they Created?
- How do I go about making these Control Charts?
- When is the Best Time to Start?
- Steps Involved in Using Statistical Process Control
- Specific SPC Tools and Procedures
- Identification and Data Gathering
- Prioritizing
- Pareto Charts
- Cause-and-Effect or Fishbone Diagram
- Flowcharting
- Scatter Plots
- Box Plot
- Control Charts - Fundamental Concepts and Key Terms
- Analysis of Quality Control Materials
- Analysis of Spiked (Fortified) Samples
- Introduction to Constructing and Interpreting Control Charts
- Shewhart Charts
- Interpretation of X-charts and X-bar Charts
- Range Chart
- Quiz #10