INTRODUCTION TO DATA ANALYSIS

Course Code: 1076

In this course, you will learn to use data analytics to create actionable recommendations, as well as identify and manage opportunities where data-based decisions can be used to change the way you do business.

Data Analysis is an ever-evolving discipline with lots of focus on new predictive modeling techniques coupled with rich analytical tools that keep increasing our capacity to handle big data. However, in order to chart a coherent path forward, it is necessary to understand where the discipline has come from since its inception.

The field of Business intelligence depends largely on Data analysis tools and techniques in order to inform effective decision-making. In fact, the disciplines are so intertwined that some often confuse the two. Therefore, we begin our introduction by examining the history of Business intelligence, its relationship to data analysis, and why the two are needed to help businesses deliver a complete assembly of their 'data puzzle'. This module also addresses some of the hurdles businesses face when dealing with data overload and suggests some possible solutions to the problem.

With the explosion of big data, businesses recognize there is a greater need for employing someone who is qualified to correctly analyze the data. In this module, we explore the qualifications for the data analyst as well as the analytic tools associated with the position. It is unfortunate that there is such a dearth of data analysts. With a projected shortage of 190,000 data science jobs into 2018, it is no wonder that businesses are scrambling to recruit talent.

What You’ll Learn

- Learn the terms, jargon, and impact of business intelligence and data analytics.
- Gain knowledge of the scope and application of data analysis.
- Explore ways to measure the performance of and improvement opportunities for business processes.
- Be able to describe the need for tracking and identifying the root causes of deviation or failure.
- Review the basic principles, properties, and application of Probability Theory.
- Discuss data distribution including Central Tendency, Variance, Normal
Distribution, and non-normal distributions.
• Learn about Statistical Inference and drawing conclusions about a Data Population.
• Learn about Forecasting, including introduction to simple Linear Regression analysis.
• Learn about Sample Sizes and Confidence Intervals and Limits, and how they influence the accuracy of your analysis.
• Explore different methods and easy algorithms for forecasting future results and to reduce current and future risk.

Who Needs to Attend
Anyone involved in operations, project management, business analysis, or management who needs an introduction to Data Analysis, would benefit from this class:
• Business Analyst, Business Systems Analyst, Staff Analyst
• Those interested in CBAP®, CCBA®, or other business analysis certifications
• Systems, Operations Research, Marketing, and other Analysts
• Project Manager, Team Leads, Project Leads, Project Assistants, Project Coordinators
• Those interested in PMP®, CAPM®, or other project management certifications
• Program Managers, Portfolio Managers, Project Management Office (PMO) staff
• Data Modelers and Administrators, DBAs
• Technical & other Subject Matter Experts (SMEs)
• IT Staff, Manager, VPs
• Finance Staff, Manager
• Operations Analyst, Supervisor
• External and Internal Consultants
• Risk Managers, Operations Risk Professionals
• Operations Managers, Line Managers, Operations Staff
• Process Improvement, Compliance, Audit, & other Governance Staff
• Thought Leaders, Transformation & Change Champions, Change Manager
• Executives, Directors, & other senior staff exploring cost reduction and process improvement options
• Executive and Administrative Assistants and Coordinators
• Job seekers and those who want to show dedication to data analysis and process improvement
• Leaders at all levels who wish to increase their Data Analysis capabilities
INTRODUCTION TO DATA ANALYSIS

Course Code: 1076

Classroom Live Outline

Part 1: What are BI and DA?
- Definitions of BI
- History of BI
- How is BI used to help Businesses
- Definition of DA
- The relationship between BI and DA

Part 2: Data Here, There, and Everywhere!
- Oracle study on business data preparedness
- Overview of Study Findings-overwhelmed by volume of data and inability to utilize data effectively
- Possible solutions to data overflow problems

Part 3: Got Data? The Unique Role of the Data Analyst
- Role of a Data Analyst
- Skill set required to be an effective Data Analyst
- Exercise: “Channeling Your Inner Analyst”
  Students are told to imagine receiving a memo from their supervisor explaining that the company is downsizing. They are expected to take on additional responsibilities including doing data analysis. They must rewrite their current job description to include the new data analyst duties.
- FACTS or Feelings: Your Choice.
  As data becomes more widely available, businesses are finding more success in adopting a fact-based decision model rather than relying on traditional intuition alone. In this module, we examine more closely the two types of decision models businesses use as well as the benefits of the fact-based model. We cover the steps of the Rational Decision Model, a fact-based method for decision making.

Part 4: Fact-Based Decision-Making Process
• The two types of Decision Models Businesses use
• The Benefits of Fact-Based Decision Making
• Rational Decision Model: Six-Step Method
• Pal’s Diner: An Example of how the Rational Model is used in practice
• Exercise: "Who’s The Boss?"
  Students are divided into groups and told to imagine that they are the CEO of their own company. They define a business-related decision that they need to make and then apply the steps of the Rational Decision Model to arrive at the final conclusion.

• 'BIG DATA' ANATOMY
  In this module, we revisit the Big Data trend with a more detailed focus. We begin by defining the buzzword—"BIG DATA", examining its core attributes, and outlining the factors that contribute to data being 'big'. We explore how businesses collect structured and unstructured data, and the challenges they face in storing and effectively using both types of data.

Part 5: Big Data Anatomy
• The Attributes of Big Data
  † Definition of Big Data
  † The 4 V’s of Big Data
  † Structured versus Unstructured Data
  † The Challenges of Big Data
• Exercise: "Camp WoeBeData"
  Students are asked to describe some of the big data challenges that their companies face and to outline what steps are being taken to address the problems.

• GETTING TO KNOW YOUR DATA
  In order to better understand how to analyze data, we first have to comprehend its depth. This requires drilling deep beneath the server it is located on and understanding its composition. Assume we are given a structured data set with labeled columns and completed rows. There are plenty of ways to summarize the story behind the data, but we cannot dive in without first getting to understand its fundamental structure. We begin by classifying the collected data as quantitative or qualitative. Then we further classify our column variables according to the way data is measured: nominal, ordinal, interval, or ratio. It is only after understanding this classification that we are able to proceed to the next step of choosing the appropriate analysis techniques which correspond to nominal, ordinal, interval or ratio variables.

Part 6: Getting to Know Your Data
• Data Types: Qualitative versus Quantitative
• Taking a Closer Look: Data Measurement
• Four Types of Data Variables
  † Definition and examples of Nominal Variables: Name only
  † Definition and examples of Ordinal Variables: Order Matters
  † Definition and examples of Interval Variables
Part 7: The Fundamental Ways we use data Visualization techniques

- The five ways we use data visualization techniques

Part 8: Displaying Tabular Data in Excel
- How to create custom tables in Excel
- How to Sort/Filter tabular data
- How to create and manipulate pivot tables

Part 9: Using Charts and Graphs to Communicate Data
- How to create Pie, Column, and Line charts using Excel
- Communicating effectively using different chart types
- How to choose the correct chart to display the correct data type
- Exercise: “Table Mining”
  Students develop tables to summarize trends in a data set related to Low Rent and Section 8 housing.
- Exercise: “Charting Poverty”
  Students develop charts and graphs to summarize the poor housing epidemic in a public housing data set.

Part 10: Using Numerical Descriptives to Summarize Data
- Measures of Centrality: Mean, Median, Mode
- Format of Data Values: Grouped Discrete and Grouped Continuous
• Formulas for the Mean
  Examples: Applying 3M’s to Grouped Discrete and Grouped Continuous Data
• Measures of Spread: Standard Deviation, Range, Inter-quartile Range
  Examples: Applying Measures of Spread to Grouped Discrete and Grouped Continuous Data
• Exercise: “Faulty Wear”
  Students use mean, median, and mode to summarize information about returns from a department store.
• Exercise: “Faulty Wear: Take 2”
  Students use measures of spread to summarize the distribution of faulty garments from a department store.
• QUANTIFYING UNCERTAINTY
  Probability is not only the most important data analysis foundation, but it is by far the strongest measure that businesses can use to quantify uncertainty. There are risky decisions that businesses encounter when investing in certain stocks and taking a chance on whether its value will rise. The insurance industry calculates the most probable life expectancy for the population and bases its rates on that uncertainty. There is an abundance of examples that illustrate why understanding probability benefits the business industry, so this module is designed to expose students to a solid understanding of the topic. We use simple, easy to understand examples to introduce students to traditional and conditional probability. Then we follow up with other business probability applications that involve relative frequency and expectation.

Part 11: Probability: Quantifying Uncertainty
• Origin of Probability
• Probability: Examples of Business Applications
• The traditional definition of Probability
• Simple Computation: The TopBottomFraction Method
• How to calculate probabilities from contingency tables
• How to Calculate conditional probability from contingency tables
• Applying probability to calculate relative frequency
• Applying probability to calculate the expected value
• Using Expected Value in Decision Making
• Exercise: “Pocket Probability”
  Students practice calculating basic probabilities using the change in their pocket.
• Exercise: “Magazine Money”
  A magazine subscription service conducted a survey to study the relationship between the number of subscriptions per household and family income. Students use contingency tables to answer questions related to conditional probability.
• Exercise: “Home Sweet Home”
  Students consider an international dataset involving the different types of home dwellings located in Bradfield, England. They use information from
contingency tables to calculate regular and conditional probabilities.

- Exercise: "Peck Up Your Speed"
  Students analyze the relative frequency of workers’ typing speeds.
- Exercise: "Rent Me"
  Students calculate the expected value in order to assist the Callow Corporation in making an important decision regarding leasing computer equipment.

- NORMALITY
  If you look at the graph of your data and notice that it looks like a bell-shaped curve, it probably follows a normal distribution. The manufacturing industry, for example, measures volumes and weights of products. Data collected from these instances are normally distributed. In fact, if we take any process and calculate its mean value enough times, then its mean value is also normally distributed. This module examines the famous ‘Normal Distribution’ and how it can be used to provide useful information about the probability of certain events.

Part 12: The Normal Distribution

- Examples of Normally Distributed Data Variables
- Characteristics of the Normal Distribution
- Interpreting the Empirical Rule
- Components of the Normal Distribution: Probabilities and X values
- Using the NORMDIST function in Excel to calculate probability from a normal distribution
- Using the NORM. INV function in Excel to calculate X values related to a normal distribution
- Exercise: "Fill ‘R’ Up"
  Students help a manufacturer use the normal distribution of volumes from a cup dispenser to calculate probabilities of certain events.

- ASSOCIATION AND PREDICTION
  If we have data measurements for at least two variables, it is natural to ask if there is some relationship between the two. This module takes a look at the two strongest measures of association, correlation, and regression, and explores how to use them to quantify the relationship. It also examines how regression output can be used to predict future observations.

Part 13: Correlation and Regression

- Definition of Correlation and Regression
- The relationship between Correlation and Regression
- Correlation Coefficient: Values
- Examples of Correlation
- Interpretation of a Regression Equation
- Step-by-Step example of How to Do a Regression Analysis
- Exercise: "Paid Sickouts"
  Students use correlation and regression to help a company determine the relationship between the number of sick days employees took and the wages they earned.
<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 19 - 20, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>RESEARCH TRIANGLE PARK, NC</td>
</tr>
<tr>
<td>Apr 23 - 24, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>BOSTON, MA</td>
</tr>
<tr>
<td>May 21 - 22, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>AUSTIN, TX</td>
</tr>
<tr>
<td>Jun 25 - 26, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>ATLANTA, GA</td>
</tr>
<tr>
<td>Aug 27 - 28, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>PHILADELPHIA, PA</td>
</tr>
<tr>
<td>Sep 24 - 25, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>MINNEAPOLIS, MN</td>
</tr>
<tr>
<td>Oct 22 - 23, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>HOUSTON, TX</td>
</tr>
<tr>
<td>Nov 19 - 20, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>DULLES, VA</td>
</tr>
<tr>
<td>Dec 17 - 18, 2020</td>
<td>8:30 AM - 4:30 PM</td>
<td>DENVER, CO</td>
</tr>
</tbody>
</table>
INTRODUCTION TO DATA ANALYSIS

Course Code: 1076

VIRTUAL CLASSROOM LIVE  $1,495 USD  2 days

Virtual Classroom Live Outline

Part 1: What are BI and DA?
  • Definitions of BI
  • History of BI
  • How is BI used to help Businesses
  • Definition of DA
  • The relationship between BI and DA

Part 2: Data Here, There, and Everywhere!
  • Oracle study on business data preparedness
  • Overview of Study Findings-overswhelmed by volume of data and inability to utilize data effectively
  • Possible solutions to data overflow problems

Part 3: Got Data? The Unique Role of the Data Analyst
  • Role of a Data Analyst
  • Skill set required to be an effective Data Analyst
  • Exercise: "Channeling Your Inner Analyst"
    Students are told to imagine receiving a memo from their supervisor explaining that the company is downsizing. They are expected to take on additional responsibilities including doing data analysis. They must rewrite their current job description to include the new data analyst duties.
  • FACTS or Feelings: Your Choice.
    As data becomes more widely available, businesses are finding more success in adopting a fact-based decision model rather than relying on traditional intuition alone. In this module, we examine more closely the two types of decision models businesses use as well as the benefits of the fact-based model. We cover the steps of the Rational Decision Model, a fact-based method for decision making.

Part 4: Fact-Based Decision-Making Process
• The two types of Decision Models Businesses use
• The Benefits of Fact-Based Decision Making
• Rational Decision Model: Six-Step Method
• Pal’s Diner: An Example of how the Rational Model is used in practice
• Exercise: "Who's The Boss?"
  Students are divided into groups and told to imagine that they are the CEO of their own company. They define a business-related decision that they need to make and then apply the steps of the Rational Decision Model to arrive at the final conclusion.

• 'BIG DATA' ANATOMY
  In this module, we revisit the Big Data trend with a more detailed focus. We begin by defining the buzzword—"BIG DATA", examining its core attributes, and outlining the factors that contribute to data being 'big'. We explore how businesses collect structured and unstructured data, and the challenges they face in storing and effectively using both types of data.

Part 5: Big Data Anatomy
• The Attributes of Big Data
  ◦ Definition of Big Data
  ◦ The 4 V's of Big Data
  ◦ Structured versus Unstructured Data
  ◦ The Challenges of Big Data
• Exercise: "Camp WoeBeData"
  Students are asked to describe some of the big data challenges that their companies face and to outline what steps are being taken to address the problems.

• GETTING TO KNOW YOUR DATA
  In order to better understand how to analyze data, we first have to comprehend its depth. This requires drilling deep beneath the server it is located on and understanding its composition. Assume we are given a structured data set with labeled columns and completed rows. There are plenty of ways to summarize the story behind the data, but we cannot dive in without first getting to understand its fundamental structure. We begin by classifying the collected data as quantitative or qualitative. Then we further classify our column variables according to the way data is measured: nominal, ordinal, interval, or ratio. It is only after understanding this classification that we are able to proceed to the next step of choosing the appropriate analysis techniques which correspond to nominal, ordinal, interval or ratio variables.

Part 6: Getting to Know Your Data
• Data Types: Qualitative versus Quantitative
• Taking a Closer Look: Data Measurement
• Four Types of Data Variables
  ◦ Definition and examples of Nominal Variables: Name only
  ◦ Definition and examples of Ordinal Variables: Order Matters
  ◦ Definition and examples of Interval Variables
Definition and examples of Ratio Variables
Summary of Statistics/Operations that can be performed on each type

Exercise: “Marketing to Low Renters”
Students are told to put on their data analyst thinking caps. They have been employed as a junior data analyst for a Marketing Company whose goal is to make a marketing campaign for a client who plans on targeting the ‘needy’ population. Students are given a public housing data set and told to classify each variable according to its measurement.

DATA VISUALIZATION
A picture is worth a thousand words, and there definitely is no exception when it comes to summarizing data. This module is dedicated to highlighting the importance of visualizing data, and how the human eye depends on visual representation to get a quick sense of data relevance. Visual representation is the audience’s first impression of the data and forms a crucial step in inviting and maintaining a genuine interest in a subject matter. We demonstrate how to create colorful, easy to understand tables, charts, and graphs that aid in helping us convey the story behind the data set being analyzed.

Part 7: The Fundamental Ways we use data Visualization techniques

Part 8: Displaying Tabular Data in Excel
• How to create custom tables in Excel
• How to Sort/Filter tabular data
• How to create and manipulate pivot tables

Part 9: Using Charts and Graphs to Communicate Data
• How to create Pie, Column, and Line charts using Excel
• Communicating effectively using different chart types
• How to choose the correct chart to display the correct data type
• Exercise: "Table Mining"
Students develop tables to summarize trends in a data set related to Low Rent and Section 8 housing.
• Exercise: "Charting Poverty"
Students develop charts and graphs to summarize the poor housing epidemic in a public housing data set.

NUMERICAL DATA SUMMARIES
Another way that data analysts summarize data is by providing a single number, or summary statistic, that has meaning. This module explores how the mean, median, and mode can be used to summarize the center of discrete and continuous grouped data. The range, standard deviation, and inter-quartile range measure the dispersion in the data set and provide information about how data points are spread.

Part 10: Using Numerical Descriptives to Summarize Data
• Measures of Centrality: Mean, Median, Mode
• Format of Data Values: Grouped Discrete and Grouped Continuous
Formulas for the Mean
 Examples: Applying 3M’s to Grouped Discrete and Grouped Continuous Data

Measures of Spread: Standard Deviation, Range, Inter-quartile Range
 Examples: Applying Measures of Spread to Grouped Discrete and Grouped Continuous Data

Exercise: “Faulty Wear”
 Students use mean, median, and mode to summarize information about returns from a department store.

Exercise: “Faulty Wear: Take 2”
 Students use measures of spread to summarize the distribution of faulty garments from a department store.

QUANTIFYING UNCERTAINTY
 Probability is not only the most important data analysis foundation, but it is by far the strongest measure that businesses can use to quantify uncertainty. There are risky decisions that businesses encounter when investing in certain stocks and taking a chance on whether its value will rise. The insurance industry calculates the most probable life expectancy for the population and bases its rates on that uncertainty. There is an abundance of examples that illustrate why understanding probability benefits the business industry, so this module is designed to expose students to a solid understanding of the topic. We use simple, easy to understand examples to introduce students to traditional and conditional probability. Then we follow up with other business probability applications that involve relative frequency and expectation.

Part 11: Probability: Quantifying Uncertainty

 Origin of Probability
 Probability: Examples of Business Applications
 The traditional definition of Probability
 Simple Computation: The TopBottomFraction Method
 How to calculate probabilities from contingency tables
 How to Calculate conditional probability from contingency tables
 Applying probability to calculate relative frequency
 Applying probability to calculate the expected value
 Using Expected Value in Decision Making
 Exercise: ”Pocket Probability”
 Students practice calculating basic probabilities using the change in their pocket.
 Exercise: ”Magazine Money”
 A magazine subscription service conducted a survey to study the relationship between the number of subscriptions per household and family income. Students use contingency tables to answer questions related to conditional probability.
 Exercise: “Home Sweet Home”
 Students consider an international dataset involving the different types of home dwellings located in Bradfield, England. They use information from
contingency tables to calculate regular and conditional probabilities.

• Exercise: “Peck Up Your Speed”
  Students analyze the relative frequency of workers’ typing speeds.

• Exercise: “Rent Me”
  Students calculate the expected value in order to assist the Callow Corporation in making an important decision regarding leasing computer equipment.

• NORMALITY
  If you look at the graph of your data and notice that it looks like a bell-shaped curve, it probably follows a normal distribution. The manufacturing industry, for example, measures volumes and weights of products. Data collected from these instances are normally distributed. In fact, if we take any process and calculate its mean value enough times, then its mean value is also normally distributed. This module examines the famous ‘Normal Distribution’ and how it can be used to provide useful information about the probability of certain events.

Part 12: The Normal Distribution

• Examples of Normally Distributed Data Variables
• Characteristics of the Normal Distribution
• Interpreting the Empirical Rule
• Components of the Normal Distribution: Probabilities and X values
• Using the NORMDIST function in Excel to calculate probability from a normal distribution
• Using the NORM. INV function in Excel to calculate X values related to a normal distribution
• Exercise: “Fill ‘R’ Up”
  Students help a manufacturer use the normal distribution of volumes from a cup dispenser to calculate probabilities of certain events.

• ASSOCIATION AND PREDICTION
  If we have data measurements for at least two variables, it is natural to ask if there is some relationship between the two. This module takes a look at the two strongest measures of association, correlation, and regression, and explores how to use them to quantify the relationship. It also examines how regression output can be used to predict future observations.

Part 13: Correlation and Regression

• Definition of Correlation and Regression
• The relationship between Correlation and Regression
• Correlation Coefficient: Values
• Examples of Correlation
• Interpretation of a Regression Equation
• Step-by-Step example of How to Do a Regression Analysis
• Exercise: “Paid Sickouts”
  Students use correlation and regression to help a company determine the relationship between the number of sick days employees took and the wages they earned.
<table>
<thead>
<tr>
<th>Date Range</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 19 - 20, 2020</td>
<td>8:30 AM - 4:30 PM EST</td>
</tr>
<tr>
<td>Apr 23 - 24, 2020</td>
<td>8:30 AM - 4:30 PM EST</td>
</tr>
<tr>
<td>May 21 - 22, 2020</td>
<td>8:30 AM - 5:30 PM EST</td>
</tr>
<tr>
<td>Jun 25 - 26, 2020</td>
<td>8:30 AM - 4:30 PM EST</td>
</tr>
<tr>
<td>Jul 23 - 24, 2020</td>
<td>11:30 AM - 7:30 PM EST</td>
</tr>
<tr>
<td>Aug 27 - 28, 2020</td>
<td>8:30 AM - 4:30 PM EST</td>
</tr>
<tr>
<td>Sep 24 - 25, 2020</td>
<td>9:30 AM - 5:30 PM EST</td>
</tr>
<tr>
<td>Oct 22 - 23, 2020</td>
<td>9:30 AM - 5:30 PM EST</td>
</tr>
<tr>
<td>Nov 19 - 20, 2020</td>
<td>8:30 AM - 4:30 PM EST</td>
</tr>
<tr>
<td>Dec 17 - 18, 2020</td>
<td>10:30 AM - 6:30 PM EST</td>
</tr>
</tbody>
</table>
INTRODUCTION TO DATA ANALYSIS
Course Code: 1076

PRIVATE GROUP TRAINING  2 days

Visit us at www.globalknowledge.com or call us at 1-866-716-6688.

Date created: 3/17/2020 3:30:16 AM
Copyright © 2020 Global Knowledge Training LLC. All Rights Reserved.