

## Syllabus

### Course Overview

The course's objective is to provide you with both the theory and practices of descriptive and inferential statistics and business analytics with data mining. This overview should enable you to apply what you learn here to your current or future business problems at work. It is important to understand that the results obtained through descriptive and inferential statistics and business analytics are only as good as the data on which the analysis is based.

Data is the main ingredient for any BI, data science, and business analytics initiative. In fact, it can be viewed as the raw material for what these popular decision technologies produce—information, insight, and knowledge (Sharda, 2017).

Traditionally, statistics is the science of collecting, integrating, cleansing, storing, describing, processing, and analyzing the data. In that last step, we create a mathematical model that best fits the data to produce a predefined output. This is data mining. In analytical terms, we use data to add new knowledge and insights that may have business value, or to spot trends that may be valuable to an organization's strategy.

Tuffery (2011) differentiates data mining as more complex from elementary descriptive statistics. As business analysts, you may hear many different terms when it comes to data mining depending on the training and experience of the individual doing the data mining. Common terms used include *concepts, cases, objects, attributes, characters, states, and variables*. *Concepts, cases, and objects* are synonyms. *Variables, attributes, characters, and states* are synonyms. In this course, we will use the terms *cases* and *variables*, though in your readings you may come across the use of other versions of these terms as listed above.

Another important point of clarification as you begin this course deals with variables. There are two types of variables—categorical (C) and quantitative (Q). Categorical variables divide cases into one or more groups. Quantitative variables are numeric and are additive. Data mining tends to focus on group cases, whereas statistics tends to focus mostly on quantitative summarization. Both methods and variables are important to the business analyst.

A typical request made of a business analyst is to understand the definition of a problem, review the problem, and then look at the data and information that may be useful in finding the best answer to the problem. However, it is often the case that the business analyst must first make sense of the data for others to understand. In this course, we will explore how an analyst could do just that.

#### Reference

Tuffery S. (2011). *Data mining and statistics for decision making*. Hoboken, NJ: John Wiley & Co.

Sharda, R. (2017). *Business intelligence, analytics, and data science: A managerial perspective* (4th ed.). Upper Saddle River, NJ: Pearson.

### Course Competencies

(Read Only)

To successfully complete this course, you will be expected to:

- 1 Evaluate the process and workflows used in data interpretation and analysis.
- 2 Evaluate the quality and constraints of source data used to solve business problems.
- 3 Apply statistical analysis and calculations in data mining.
- 4 Use statistical software to manipulate data to solve business problems.
- 5 Communicate effectively.

### Course Prerequisites

Prerequisite(s): BUS4220.

## Syllabus >> Course Materials

### Required

The materials listed below are required to complete the learning activities in this course.

### Integrated Materials

Many of your required books are available via the VitalSource Bookshelf link in the courseroom, located in your Course Tools. Registered learners in a Resource Kit program can access these materials using the courseroom link on the Friday before the course start date. Some materials are available only in hard-copy format or by using an access code. For these materials, you will receive an email with further instructions for access. Visit the [Course Materials](#) page on Campus for more information.

#### Book

Sharda, R., Delen, D., & Turban, E. (2018). *Business intelligence, analytics, and data science: A managerial perspective* (4th ed.). New York, NY: Pearson. ISBN: 9780134633282.

### Library

The following required readings are provided in the Capella University Library or linked directly in this course. To find specific readings by journal or book title, use [Journal and Book Locator](#). Refer to the [Journal and Book Locator library guide](#) to learn how to use this tool.

- Ahlemeyer-Stubbe, A., & Coleman, S. (2014). *A practical guide to data mining for business and industry*. Hoboken, NJ: Wiley.
- Alnougari, M., & Hanano, A. (2017). *Integration of business intelligence with corporate strategic management*. *Journal of Intelligence Studies in Business*, 7(2), 5–16.
- Bedeian, A. G. (2014). "More than meets the eye": A guide to interpreting the descriptive statistics and correlation matrices reported in management research. *Academy of Management Learning and Education*, 13(1), 121–135.
- Bāliņa, S., Žuka, R., & Krasts, J. (2016). Opportunities for the use of business data analysis technologies. *Economics and Business*, 28, 20–25.
- Campbell, J. (2017). *Exploring machine learning: k-nearest neighbor (kNN)* [Video]. Skillssoft Ireland Limited.
- Campbell, J. (2017). *Exploring machine learning: k-nearest neighbor with scikit-learn* [Video]. Skillssoft Ireland Limited.
- Kotsiantis, S. B. (2013). *Decision trees: A recent overview*. *The Artificial Intelligence Review*, 39(4), 261–283.
- Lachance, D. (2016). *Microsoft power BI: Business intelligence overview* [Video]. Skillssoft Ireland Limited.
- Lachance, D. (2016). *Microsoft power BI: Self-service business intelligence* [Video]. Skillssoft Ireland Limited.
- Maheshwari, A. (2014). *Business intelligence and data mining*. Business Expert Press.

- Miller, W. (2016). *Python for data science: Data preprocessing and text mining* [Video]. Skillssoft Ireland Limited.
- Obeidat, M., North, M., Richardson, R., Rattanak, V., & North, S. (2015). *Business intelligence technology, applications, and trends*. *International Management Review*, 11(2), 47–56, 113.
- Robinson, C. (2013). *Big data: Decision trees* [Video]. Skillssoft Ireland Limited.
- Scott, S. (2015). *Data science fundamentals in R: k-nearest neighbor classification* [Video]. Skillssoft Ireland Limited.
- Shmueli, G., Bruce, P. C., & Patel, N. R. (2016). *Data mining for business analytics: Concepts, techniques, and applications with XLMiner* (3rd ed.). Hoboken, NJ: Wiley.
- Sidorski, A. (2018). *Implementing ML algorithm using scikit-learn: Decision tree classification* [Video]. Skillssoft Ireland Limited.
- Sidorski, A. (2018). *Supervised, unsupervised & deep learning: Text mining and data assembly* [Video]. Skillssoft Ireland Limited.
- Strickland, J. (2016). *Predictive analytics: Data mining* [Video]. Skillssoft Ireland.
- Strickland, J. (2016). *Predictive analytics: Decision trees* [Video]. Skillssoft Ireland Limited.
- Theriault, P. (2015). *Office 2016 first look: Discovering the business intelligence features* [Video]. Skillssoft Ireland Limited.
- Wills, M. J. (2014). Decisions through data: Analytics in healthcare. *Journal of Healthcare Management*, 59(4), 254–262.
- Xia, B. S., & Gong, P. (2014). *Review of business intelligence through data analysis*. *Benchmarking*, 21(2), 300–311.

### External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- SAS. (n.d.). [SAS OnDemand for academics](https://odamid.oda.sas.com/SASODARegistration/). Retrieved from <https://odamid.oda.sas.com/SASODARegistration/>
- Tableau. (n.d.). [Tableau for students](https://www.tableau.com/academic/students). Retrieved from <https://www.tableau.com/academic/students>

### Suggested

The following materials are recommended to provide you with a better understanding of the topics in this course. These materials are not required to complete the course, but they are aligned to course activities and assessments and are highly recommended for your use.

### Optional

The following optional materials are offered to provide you with a better understanding of the topics in this course. These materials are not required to complete the course.

### Library

The following optional readings may be available in the Capella University Library. To find specific readings by journal or book title, use [Journal and Book Locator](#). Refer to the [Journal and Book Locator library guide](#) to learn how to use this tool. If the full text is not available, you may be able to request a copy through the [Interlibrary Loan](#) service.

- Thuraisingham, B. (2003). *Web data mining and applications in business intelligence and counter-terrorism*. Boca Raton, FL: CRC Press.

### External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- Front Line Systems, Inc. (2018) [Analytic solver data mining](https://www.solver.com/xlminer-platform). Retrieved from <https://www.solver.com/xlminer-platform>
- Optional software: Analytic Solver, Qlik Sense Desktop, SAS Enterprise Guide, SAS Enterprise Miner, or Tableau Desktop. Directions for downloading the software are provided in the Unit 1 studies.
- Qlik. (n.d.). [Getting started using Qlik Sense](https://help.qlik.com/en-US/sense/September2018/Subsystems/Hub/Content/Sense_Hub/Introduction/get-started.htm). Retrieved from [https://help.qlik.com/en-US/sense/September2018/Subsystems/Hub/Content/Sense\\_Hub/Introduction/get-started.htm](https://help.qlik.com/en-US/sense/September2018/Subsystems/Hub/Content/Sense_Hub/Introduction/get-started.htm)
- Qlik. (n.d.). [Try Qlik Sense Desktop for free](https://www.qlik.com/us/try-or-buy/download-qlik-sense). Retrieved from <https://www.qlik.com/us/try-or-buy/download-qlik-sense>
- Tableau. (n.d.). [Tableau community: Student resource page](https://community.tableau.com/community/students/overview). Retrieved from <https://community.tableau.com/community/students/overview>

## Unit 1 >> Basic Statistics and Business Analytics: Interpretation and Analysis

### Introduction

Statistics can be defined as "the science of collecting, describing, and analyzing the data" (Lock, Lock, Lock Morgan, Lock, and Lock, 2013, p. 4) with models. Business analysts must have an understanding of this science, the data they use, and the distribution of that dataset. Foreknowledge of dataset distribution represents a difference between traditional statistics and data mining. Data interpretation is the process of assigning meaning to the result set by answering how useful the data is to the objective of the test and what it means to decision makers. In assisting with decision making, pertinent actionable knowledge produced about an objective is the Business Intelligence (BI).

Complete the [Statistics and Business Analytics: Interpretation and Analysis](#) multimedia interactive to learn more.

#### Reference

Lock, R. H., Lock, P. F., Lock Morgan, K., Lock, E. F., & Lock, D. F. (2013). *Statistics: Unlocking the power of data*. Hoboken, NJ: John Wiley & Sons, Inc.

## Learning Activities

### u01s1 - Studies

## Readings

Use the Capella University Library to read the following:

- In your [A Practical Guide to Data Mining for Business and Industry](#) text:
  - Part II: Section 3, "All About Data," pages 33–59.
- Alhoukari, M., & Hanano, A. (2017). [Integration of business intelligence with corporate strategic management](#). *Journal of Intelligence Studies in Business*, 7(2), 5–16.
- Wills, M. J. (2014). [Decisions through data: Analytics in healthcare](#). *Journal of Healthcare Management*, 59(4), 254–262.

## Capella Resource

To learn more about statistics, see [Understanding Statistics](#) on Campus.

## Multimedia

Complete the following Capella multimedia presentation:

- [Statistics and Business Analytics: Interpretation and Analysis](#).

## Videos

Complete the following brief instructional videos.

- Theriault, P. (2015). [Office 2016 first look: Discovering the business intelligence features \[Video\]](#). Skillssoft Ireland Limited.
- Lachance, D. (2016). [Microsoft power BI: Business intelligence overview \[Video\]](#). Skillssoft Ireland Limited.
- Lachance, D. (2016). [Microsoft power BI: Self-service business intelligence \[Video\]](#). Skillssoft Ireland Limited.

### u01s1 - Learning Components

- Explain the benefits of central tendency measures.
- Describe the limitations of central tendency measures.
- Give examples of statistical methods for descriptive statistics and business analytics.
- Explain the importance of understanding the objectives and outcomes of the business problem before developing a statistic model or data mining model.
- Describe how source constraints, like a survey, can affect the outcome of an analytical project.

### u01s2 - Software Preparation and Technology Access

In this course, you will be using software and technology that is needed to complete designated activities and assignments. There is no additional cost for this software and technology. Some software packages will be made available to you at no additional cost through Capella's subscription with Microsoft, while other software packages are available for free download through open-source licensing.

Capella University requires learners to meet certain minimum [computer requirements](#). Please note that some software required for a course may exceed these minimum requirements. Check the requirements for the software you may need to download and install to make sure it will work on your device. Most software will require a Windows PC. If you use a Mac, refer to [Installing a Virtual Environment and Windows on a Mac](#).

The software and technologies below are strongly recommended to support you in completing the course objectives. If you have access to other tools that you believe may still meet the requirements of this course, please discuss your selected alternatives with your instructor.

If you use assistive technology or any alternative communication methods to access course content, please contact [Disability Services](#) with any access-related questions or to request accommodations.

For this course, follow the instructions provided through the links below to download and install software or register for an account, as required. If you encounter any difficulties in the download and installation process, post a detailed question in the Ask Your Instructor section of the course. Your instructor should be able to help you or point you in the right direction for the answers you need.

## SAS Software

SAS software is optional for this course. You may choose to use it for assignments in this course where appropriate, but it is not required.

SAS Enterprise Guide and SAS Enterprise Miner are available for you to use during this course for your statistical calculations; there is no fee for you to use these resources. SAS is one of the most commonly used statistical analysis tools in business and, as a Capella learner, you have access to this valuable resource.

To gain free access to this software, follow the steps at the following site and the campus resource:

- SAS. (n.d.). [SAS OnDemand for academics](https://odamid.oda.sas.com/SASODARegistration/). Retrieved from <https://odamid.oda.sas.com/SASODARegistration/>
- [SAS OnDemand for Academics \(SODA\)](#).
  - Use the [Importing Excel Data Using SAS \[PDF\]](#) tutorial to learn the steps required when importing Excel data in SAS. Note that the variables titles in your file should always occupy row one.

## Tableau Desktop Software

As a Capella learner, you are eligible for a free one-year Tableau Desktop license. You may choose to use it for assignments in this course where appropriate, but it is not required. To request a license:

1. Navigate to the following site to download the free registration form:
  - Tableau. (n.d.). [Tableau for students](https://www.tableau.com/academic/students). Retrieved from <https://www.tableau.com/academic/students>
2. Fill out the registration form.
3. Follow the remaining directions as they are presented to you to download and install the software.

Tutorials and training videos, user guides and other support is available from the following page:

- Tableau. (n.d.). [Tableau community: Student resource page](https://community.tableau.com/community/students/overview). Retrieved from <https://community.tableau.com/community/students/overview>

## Qlik Sense® Desktop Software

As a Capella learner, you are eligible for a free one-year Qlik Sense Desktop license. You can choose to use it for assignments in this course where appropriate, but it is not required. To register for a Qlik account and download Qlik Sense Desktop:

1. Navigate to the following page:
  - Qlik. (n.d.). [Try Qlik Sense Desktop for free](https://www.qlik.com/us/try-or-buy/download-qlik-sense). Retrieved from <https://www.qlik.com/us/try-or-buy/download-qlik-sense>
2. Fill out the registration form found on the page.
3. Follow the remaining directions as they are presented to you to download and install the software.

Tutorials and training videos, user guides and other support is available from the following page:

- Qlik. (n.d.). [Getting started using Qlik Sense](https://help.qlik.com/en-US/sense/September2018/Subsystems/Hub/Content/Sense_Hub/Introduction/get-started.htm). Retrieved from [https://help.qlik.com/en-US/sense/September2018/Subsystems/Hub/Content/Sense\\_Hub/Introduction/get-started.htm](https://help.qlik.com/en-US/sense/September2018/Subsystems/Hub/Content/Sense_Hub/Introduction/get-started.htm)
  - Select the release month and year on the left side of the screen to access resources for the release you downloaded.

## Analytic Solver Software

As a Capella learner, you are eligible for a 15-day free trial license for the Analytic Solver data mining add-In for Excel (formerly XLMiner). You can choose to use it for assignments in this course where appropriate, but it is not required. To download the software:

1. Navigate to the following page on the FrontlineSolvers website:
  - Front Line Systems, Inc. (2018) [Analytic solver data mining](https://www.solver.com/xlminer-platform). Retrieved from <https://www.solver.com/xlminer-platform>

2. Fill out and submit the registration form found on the page.
3. Follow the remaining directions as they are presented to you to download and install the software. Your free trial includes a full user guide and reference guide.

## Additional Online Resources

As a Capella learner, you have access to IT online resources through Capella's [Skillsoft](#) subscription, where you can find helpful materials.

### u01a1 - What's for Dessert?

By successfully completing this assignment, you will demonstrate your proficiency in the following course competencies and assignment criteria:

- Competency 1: Evaluate the process and workflows used in data interpretation and analysis.
  - Explain the benefits of central tendency measures.
  - Describe the limitations of central tendency measures.
  - Describe how source constraints, like a survey, can affect the outcome of an analytical project.
- Competency 2: Evaluate the quality and constraints of source data used to solve business problems.
  - Explain the importance of understanding the objectives and outcomes of the business problem before developing a statistic model or data mining model.
- Competency 3: Apply statistical analysis and calculations in data mining.
  - Give examples of statistical methods for descriptive statistics and business analytics.
- Competency 5: Communicate effectively.
  - Communicate effectively.

## Instructions

What is the best item to place on the dessert menu? There are many factors to consider when developing and selling a product. Deciding what to place on a dessert menu is an example problem.

Consider this scenario:

A general manager (GM) of a local restaurant is considering changing the dessert items on the current menu. One factor the GM faces is the dietary concerns of the restaurant's patrons: lactose intolerance, sugar intake concerns, and food allergies, just to name a few. The GM also needs to consider the restaurant's expertise in making desserts. The GM determines that the restaurant's bakers and chefs are skilled in making frozen strawberry yogurt, creamy orange ice cream, key lime pie, and four-layer chocolate cakes.

You are a local business intelligence consultant who specializes in hospitality. The GM contacts you to determine which dessert item to place on the menu for a given day of the week and to include in special events (Administrative Professionals Day, Mother's Day, Father's Day, Boss's Day, and so on). The GM identifies the following questions and a couple of requests for a survey instrument that the restaurant's patrons will complete:

- What percentage of patrons have a dietary concern?
- What percentage of patrons are interested in trying at least one of the dessert options?
- Does the patron eat dessert?
- Which of the dessert options is most appealing to the patron?
  - Rank the four dessert types with four (4) being the most appealing and one (1) being very unappealing.
  - Rate each of the four dessert types on a 1–10 scale with ten being extremely appealing and one being very unappealing.
- How do the ratings change between days of the week?
- How often do patrons visit the restaurant per month?
- On which days do patrons visit the restaurant?

Create a model from the data provided in the Just Desserts Data workbook, linked in the Resources. Then provide a 2–3-page report sharing your model, analysis, and recommendations, and listing the number of possible variables for a descriptive model. Be sure to identify the cases for the data the GM is collecting. Describe the variable types that should be included in the dataset, and describe at least one question in addition to those listed above that you think would help the GM.

We used descriptive statistics to come up with a recommendation. Is the Business Intelligence generated sufficient for decision making? If yes, explain why it is sufficient. If no, explain the reason and propose a solution.

Be sure to address each of the scoring guide criteria.

## Other Requirements

- **Written communication:** Written communication should be free of errors that detract from the overall message.
- **APA formatting:** Your paper should be formatted according to the current APA style and formatting and should include a cover page and table of contents.
- **Length:** 2–3 typed and double-spaced pages, plus a cover page and table of contents.
- **Font and font size:** Times New Roman, 12 point.

Course Resources

---

Just Desserts Data [XLSX]

---

[APA Style and Format](#)

### u01d1 - Population and Sample

Discuss how a population and a sample of data differ. Why are central tendency measures helpful when performing the analysis of a dataset? What is a limitation of central tendency when using descriptive statistics? Are the measures of central tendency good enough to describe the major properties of a dataset? Does the measures of variance value adding in describing a dataset? Why? Are the measures of central tendencies and variances sufficient enough to provide actionable Business Intelligence?

## Response Guidelines

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

Course Resources

---

Undergraduate Discussion Participation Scoring Guide

### u01d1 - Learning Components

- Differentiate a population and a sample of data.
- Discuss measures of central tendencies and variances.

### u01d2 - Recommendations to General Manager

Share your findings from this unit's assignment and your recommendation for the general manager (GM). What did you find interesting? What other techniques to obtain a viable dataset could be used to help the GM? What are some limitations of surveys? Assume that in using this dataset, you have calculated an estimate for an important sales attribute, such as internal rate of return (IRR). What if we had also calculated a confidence level? Would this help in decision making? How? Do you want to add another BI model? Which one?

## Response Guidelines

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

Course Resources

---

## Undergraduate Discussion Participation Scoring Guide

### u01d2 - Learning Components

- Discuss techniques to obtain a viable dataset and make decisions.

## Unit 2 >> Decision Trees – An Introduction

### Introduction

Traditional decision trees can help decision makers decide on an action by representing the different critical points graphically (Ragsdale, 2015). According to Larose and Larose (2015), a decision tree is one of the attractive classification methods. A decision tree is a collection of nodes with branches connecting the nodes to each other. There are three types of nodes—a decision or square node, event or circular node, and terminal node or leaves. Decision nodes represent the question or decision at hand. Event nodes represent uncertain outcomes (risks) to an event. Terminal nodes, small solid dots at the end of event branches, indicate a possible outcome.

Decision trees can be used in a wide variety of business decisions by helping business analysts to understand and make sense of available data, create actionable Business Intelligence, and to make decisions through forecasting. Still, it is important to understand that even the best decision-tree model does not guarantee the projected outcome. In a decision tree, the theory of probability comes into play. Like most models, a decision tree provides a range of possible outcomes with the likelihood based on inputs. This is why testing and refactoring of even the most basic business intelligence models by business analysts is so important.

Complete the [Decision Trees](#) multimedia interactive to learn more.

### References

Larose, T. D., & Larose, C. D. (2015). *Data mining and predictive analytics*. Hoboken, NJ: Wiley.

Ragsdale, C. (2015). *Spreadsheet modeling and decision analysis: A practical introduction to management science* (6th ed.). Mason, OH: Cengage Learning.

### Learning Activities

#### u02s1 - Studies

## Readings

Use your *Business Intelligence, Analytics and Data Science: A Managerial Perspective* text to read the following:

- Section 6.8, "Decision Analysis with Decision Tables and Decision Trees," pages 349–351.

Use the Capella University Library to read the following:

- In your *A Practical Guide to Data Mining for Business and Industry* text:
  - Part II, Section 6.5, "Decision Trees," pages 127–136.
  - Part II, Section 6.6, "Neural Networks," page 137–140.
  - Part II, Section 6.7, "Which Method Produces the Best Model? A Comparison of Regression, Decision Trees, and Neural Networks," page 141.
- Kotsiantis, S. B. (2013). [Decision trees: A recent overview](#). *The Artificial Intelligence Review*, 39(4), 261–283.
- Obeidat, M., North, M., Richardson, R., Rattanak, V., & North, S. (2015). [Business intelligence technology applications and trends](#). *International Management Review*, 11(2), 47–56, 113.

## Multimedia

Complete the following Capella multimedia presentation:

- [Decision Trees](#).

## Videos

Complete the following brief instructional videos.

- Strickland, J. (2016). [Predictive analytics: Decision trees \[Video\]](#). Skillssoft Ireland Limited.

- Robinson, C. (2013). *Big data: Decision trees* [Video]. Skillssoft Ireland Limited.
- Sidorski, A. (2018). *Implementing ML algorithm using scikit-learn: Decision tree classification* [Video]. Skillssoft Ireland Limited.

#### u02s1 - Learning Components

- Explain the strengths and weaknesses of a decision-tree model.
- Explain how traditional decision trees help business analysts recommend a course of action.
- Create a traditional decision-tree model using software.

#### u02s2 - Optional: Software Download and Installation

You may download and use any of following software tools, as appropriate, for assignments in this course.

- SAS Enterprise Guide.
- SAS Enterprise Miner.
- Tableau Desktop.
- Qlik Sense Desktop.
- Analytic Solver (formerly XLMiner).

Directions for downloading the software are provided in the Unit 1 studies.

#### u02a1 - Decision Trees - Actions to Take

By successfully completing this assignment, you will demonstrate your proficiency in the following course competencies and assignment criteria:

- Competency 2: Evaluate the quality and constraints of source data used to solve business problems.
  - Explain the strengths and weaknesses of a decision-tree model.
- Competency 3: Apply statistical analysis and calculations in data mining.
  - Explain how traditional decision trees help business analysts recommend a course of action.
- Competency 4: Use statistical software to manipulate data to solve business problems.
  - Create a traditional decision-tree model using software.
- Competency 5: Communicate effectively.
  - Communicate effectively.

## Instructions

How to help with decisions using a traditional decision tree?

Create a decision-tree model based on the business case contained in the Not Such a Sure Bet document linked in the Resources or create a decision-tree model adopted to a business case in the industry of your choice. Note that the model must be flexible to capture a change at any point and recalculate a new outcome.

After creating your decision-tree model, explain the outcome you reached through your model, and make a recommendation for action based on it if you have enough actionable Business Intelligence. If you do not have enough actionable Business Intelligence, what alternative methodology can you use to generate it and how will you use it? Justify your recommendation. Write an analysis in which you discuss the limitation of the proposed solution and its applicability to the real-life business problem. How do you compare decision tree models in the class of classification methods? Be sure to address the scoring guide criteria in your analysis.

## Other Requirements

- **Written communication:** Written communication should be free of errors that detract from the overall message.
- **APA formatting:** Your paper should be formatted according to the current APA style and formatting and should include a cover page and table of contents.
- **Length:** 2–3 typed and double-spaced pages, plus a cover page and table of contents.
- **Font and font size:** Times New Roman, 12 point.

### **u02d1 - Decision Trees - Strengths and Weaknesses**

Based on this unit's readings, what did you learn about traditional decision-tree modeling? Describe at least one strength and one weakness of decision-tree models. What is something that an analyst can do to help decision makers understand the strengths and weaknesses of a decision-tree model? Discuss the benefits and drawbacks of a binary tree versus a many-branched tree. Is the knowledge generated actionable?

### **Response Guidelines**

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

#### u02d1 - Learning Components

- Propose ways to help decision makers understand strengths and weaknesses of decision-trees.
- Discuss the benefits and drawbacks of binary versus many-branched trees.

### **u02d2 - Decision Tree Outcomes**

Explain the outcome of your decision-tree model from this unit's assignment and share the recommendation you made because of the model's outcome. Is this the decision you would have made before examining the outcome of your model? What other factors would you include in the model to help with your decision? Can you make a decision using another analytics method? Which method, and why? Which method provides you with stronger Business Intelligence for decision making?

### **Response Guidelines**

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

#### u02d2 - Learning Components

- Recommend factors to improve outcomes of decision-tree models.
- Compare methods of generating Business Intelligence for decision making.

## **Unit 3 >> Data Mining – A Review**

### **Introduction**

Not all data patterns are interesting or relevant. Data mining is a method of applying several components and processes in order to find patterns within a dataset (Han, Kamber, & Pei, 2012). Where traditional statistics knows the distribution of a dataset; in data mining scenarios, the distribution of a dataset is initially unknown.

Data mining is more than creating a model. A business analyst running a successful data mining project or program knows the data, understands the objective the data mining project is solving, and through exploration quickly decides whether the data is useful to extract information that decision-makers are required to act upon. Not all data is useful, and most patterns found in data are noise. Shmueli, Patel, and Bruce (2010) outline the steps of data mining as follows:

1. Develop and understand the purpose of the data mining project.
2. Obtain the dataset to be used in the analysis.
3. Explore, clean, and process data.
4. Reduce the data.
5. Determine the data mining task.
6. Choose the data mining technique.
7. Use an algorithm to perform the task.
8. Interpret the results of the algorithm.
9. Deploy the model.

For this unit, we will review general rules for clustering and association as an introduction to data mining. Data mining algorithms can be supervised or unsupervised. Unsupervised learning algorithms are descriptive and can help find patterns that lead to the development of a predictive model.

A machine-learning, decision-tree model meets the definition of both a descriptive and predictive tool for business analysts. Machine-learning decision trees are different from traditional decision trees in that the data decides how to construct the decision tree. This example of a supervised algorithm breaks the data into datasets that allow for analysis outside the current objectives of a given data mining project.

Complete the [Data Mining](#) multimedia interactive to learn more.

#### Reference

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. Waltham, MA: Elsevier/Morgan Kaufmann.

Shmueli, G., Patel, N. R., & Bruce, P. C. (2010). *Data mining for business intelligence*. Hoboken, NJ: Wiley.

### Learning Activities

#### u03s1 - Studies

## Readings

Use your [Data Mining for Business Analytics: Concepts, Techniques, and Applications With XLMiner](#) text to read the following:

- Chapter 4, "Dimension Reduction," pages 82–105.

Use the Capella University Library to read the following:

- In your [A Practical Guide to Data Mining for Business and Industry](#) text:
  - Part I, "Data Mining Concept," pages 3–30.
    - Read all sections, Data mining concept.
  - Review Part II: Section 3, "All About Data," pages 33–59.
- Xia, B. S., & Gong, P. (2014). [Review of business intelligence through data analysis](#). *Benchmarking*, 21(2), 300–311.
- Maheshwari, A. (2014). [Business intelligence and data mining](#). Business Expert Press.
  - Chapter 4, "Data Mining," pages 45–60.

## Multimedia

Complete the following Capella multimedia presentation:

- [Data Mining](#).

## Videos

Complete the following brief instructional videos:

- Strickland, J. (2016). *Predictive analytics: Data mining* [Video]. Skillssoft Ireland.
- Sidorski, A. (2018). *Supervised, unsupervised & deep learning: Text mining and data assembly* [Video]. Skillssoft Ireland Limited.

## Optional – Reading

Use the Capella University Library to complete the following optional reading:

- Thuraisingham, B. (2003). *Web data mining and applications in business intelligence and counter-terrorism*. Boca Raton, FL: CRC Press.

### u03s1 - Learning Components

- Describe how data mining or machine learning can help solve business problems.
- Give examples of the different data mining strategies for descriptive analytics.
- Explain some of the methods used in data mining for descriptive analytics to create actual knowledge.
- Explain how a problem definition for a descriptive analytics mining project may skew the results of a future predictive model.

### u03s2 - Optional: Software Download and Installation

You may download and use any of following software tools, as appropriate, for assignments in this course.

- SAS Enterprise Guide.
- SAS Enterprise Miner.
- Tableau Desktop.
- Qlik Sense Desktop.
- Analytic Solver (formerly XLMiner).

Directions for downloading the software are provided in the Unit 1 studies.

### u03a1 - Recommendations With Possible Bias

By successfully completing this assignment, you will demonstrate your proficiency in the following course competencies and assignment criteria:

- Competency 1: Evaluate the process and workflows used in data interpretation and analysis.
  - Explain some of the methods used in data mining for descriptive analytics to create actual knowledge.
- Competency 2: Evaluate the quality and constraints of source data used to solve business problems.
  - Explain how a problem definition for a descriptive analytics mining project may skew the results of a future predictive model.
- Competency 3: Apply statistical analysis and calculations in data mining.
  - Describe how data mining or machine learning can help solve business problems.
  - Give examples of the different data mining strategies for descriptive analytics.
- Competency 5: Communicate effectively.
  - Communicate effectively.

## Instructions

Use the following scenario to complete your assignment:

You are a business analyst for a small manufacturing company in a small town of 12,000 people along one of the Great Lakes. The company enjoys a 46 percent market share, strong profits, and happy shareholders (owners). In your last meeting, you were tasked with making a recommendation for a new product line of widgets that recreational boaters could use to enhance their boats. Money is no object, so you were only provided data by the manager of operations. The owners of the company have two objectives—grow the business by at least 18 percent for every new product and only start a project if there is at least an 85 percent probability that the new product line will sell for at least 33 percent over the cost of developing and producing the product.

Using the information provided in the Do We Go for It? document, linked in the Resources, create a decision-tree model and then craft an analysis based on your model in which you do the following:

- Recommend for or against starting a new product line, explaining how you arrived at that recommendation.

- Explain any bias that may exist within your model.
- Identify additional methods to test, enhance, and support your recommendations.
- Explain any warnings for the company's decision makers.
- Be sure to address the scoring guide criteria in your analysis and to include your model in the appendix.

## Other Requirements

- **Written communication:** Written communication should be free of errors that detract from the overall message.
- **APA formatting:** Your paper should be formatted according to the current APA style and formatting and should include a cover page and table of contents or appendix.
- **Length:** 2–3 typed and double-spaced pages, plus a cover page and table of contents.
- **Font and font size:** Times New Roman, 12 point.

Course Resources

---

Do We Go for It?

---

[APA Style and Format](#)

### u03d1 - Elements of a Data Mining Strategy

What are important factors to consider when evaluating a data mining project? Why is it important to understand the problem statement and objective before implementing a data mining project? How does the required Business Intelligence determine the methods of data mining to be applied?

## Response Guidelines

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

Course Resources

---

Undergraduate Discussion Participation Scoring Guide

u03d1 - Learning Components

- Discuss factors to consider when evaluating a data mining project.

### u03d2 - Making Sense of the Data

Describe an approach a business analyst can take to make sense of the data for decision makers. What is the difference between understanding and providing the best alternative to the problem and discovering the data story from the data? How does data mining help organizations?

## Response Guidelines

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

Course Resources

---

Undergraduate Discussion Participation Scoring Guide

u03d2 - Learning Components

- Compare approaches to make sense of data for decision makers.
- Discuss how data mining can help solve business problems.

## Unit 4 >> k-Nearest Neighbor

### Introduction

K-nearest neighbor is a supervised machine-learning, data-mining algorithm. The majority wins with a k-nearest neighbor—a democracy in action. The k-nearest neighbor algorithm is trained with some number of variables, and then, based on the training, the business analyst tests the model for various expected outcomes, generating Business Intelligence.

How does one accomplish and test a k-nearest neighbor algorithm? When new data is available to the model, the k-nearest neighbor calculates the distance from the closest data points and then votes on the most likely outcome for the new data point.

The important thing to remember about k-nearest neighbor is that the majority wins. So, say that the new data point is closest to the expected outcome three out of five times, or three to two, then the majority wins and the k-nearest neighbor algorithm will predict that the new data point is more likely to conclude with the expected outcome.

The accuracy of k-nearest neighbor is easy to assess. Just run the same number of historical records as the training data with the known outcome of the new record set. The data mining model will quickly provide a score on the accuracy of the algorithm itself.

Complete the [K-Nearest Neighbor](#) multimedia interactive to learn more.

### Learning Activities

#### u04s1 - Studies

## Readings

Use your *Data Mining for Business Analytics: Concepts, Techniques, and Applications With XLMiner* text to read the following:

- Chapter 7, "k-Nearest-Neighbors (k-NN)," pages 157–168.

Use your *Business Intelligence, Analytics and Data Science: A Managerial Perspective* text to read the following:

- Chapter 4, "Predictive Analytics I: Data Mining Process, Methods, and Algorithms," pages 190–241.

Use the Capella University Library to read the following:

- In your *A Practical Guide to Data Mining for Business and Industry* text:
  - Review Part II, Section 6.5, "Decision Trees," pages 129–136.
  - Read Part II, Section 6.9, "Cluster Analysis," pages 148–151.
  - Read Part II, Section 6.11, "Group Purchase Methods: Association and Sequence Analysis," pages 155–160.

## Multimedia

Complete the following Capella multimedia presentation:

- [K-Nearest Neighbor](#).

## Videos

Complete the following brief instructional videos.

- Campbell, J. (2017). *Exploring machine learning: k-nearest neighbor (kNN)* [Video]. Skillssoft Ireland Limited.
- Scott, S. (2015). *Data science fundamentals in R: k-nearest neighbor classification* [Video]. Skillssoft Ireland Limited.
- Campbell, J. (2017). *Exploring machine learning: k-nearest neighbor with scikit-learn* [Video]. Skillssoft Ireland Limited.

#### u04s1 - Learning Components

- Explain how k-nearest neighbor, a prescriptive data mining technique, is applied within business analytics.
- Explain where and when to use k-nearest neighbor when using the data mining technique for recommendations.
- Create a k-nearest neighbor data mining model with a statistical software tool.
- Describe how the accuracy of the test data can influence a data mining model.

#### u04s2 - Optional: Software Download and Installation

You may download and use any of following software tools, as appropriate, for assignments in this course.

- SAS Enterprise Guide.
- SAS Enterprise Miner.
- Tableau Desktop.
- Qlik Sense Desktop.
- Analytic Solver (formerly XLMiner).

Directions for downloading the software are provided in the Unit 1 studies.

#### u04a1 - k-Nearest Neighbor

By successfully completing this assignment, you will demonstrate your proficiency in the following course competencies and assignment criteria:

- Competency 1: Evaluate the process and workflows used in data interpretation and analysis.
  - Explain how k-nearest neighbor, a prescriptive data mining technique, is applied within business analytics.
- Competency 2: Evaluate the quality and constraints of source data used to solve business problems.
  - Describe how the accuracy of the test data can influence a data mining model.
- Competency 3: Apply statistical analysis and calculations in data mining.
  - Explain where and when to use k-nearest neighbor when using the data mining technique for recommendations.
- Competency 4: Use statistical software to manipulate data to solve business problems.
  - Create a k-nearest neighbor data mining model with a statistical software tool.
- Competency 5: Communicate effectively.
  - Communicate effectively.

## Instructions

Create a data mining k-nearest neighbor model using the Who Has the Most Credits? spreadsheet, linked in the Resources, as the dataset. You may use SAS Enterprise Guide, Enterprise Miner, or any other software to create your model.

Then, based on your model, write an analysis that does the following:

- Interpret the model you created in the context of Business Intelligence generation. How many customers are more likely to purchase a widget from the corporation represented in the spreadsheet? Given the distance between the training dataset and the visiting customers, what recommendations would you make to improve the model? Explain how you would model the data using a decision-tree model.
- Evaluate the company's true objective regarding data mining and Business Intelligence.

Be sure to address each of the scoring guide criteria in your analysis and to include the model you created in the appendix of your paper.

## Other Requirements

- **Written communication:** Written communication should be free of errors that detract from the overall message.
- **APA formatting:** Your paper should be formatted according to the current APA style and formatting and should include a cover page and table of contents.
- **Length:** 2–3 typed and double-spaced pages, plus a cover page, table of contents, and appendix.
- **Font and font size:** Times New Roman, 12 point.

Who Has the Most Credits? [XLSX]

---

[APA Style and Format](#)

#### **u04d1 - Traditional Decision Tree Versus K-Nearest Neighbor**

Explain the difference between a traditional decision-tree model and a k-nearest neighbor model. How might an organization use k-nearest neighbor to help with its business by generating Business Intelligence? How would the results compare with Business Intelligence generated by a decision tree?

### **Response Guidelines**

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

Undergraduate Discussion Participation Scoring Guide

u04d1 - Learning Components

- Compare decision trees and k-nearest neighbor models, including the Business Intelligence generated by each.

#### **u04d2 - Challenges of k-Nearest Neighbor**

Discuss the challenges you had creating a k-nearest neighbor by hand. What did you learn using a k-nearest neighbor approach to provide the best alternative for the assignment? Share the recommendations you made for the assignment and why you feel they could enhance the current model. Can you identify another business or real-life application for the k-nearest neighbor method?

### **Response Guidelines**

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

Undergraduate Discussion Participation Scoring Guide

u04d2 - Learning Components

- Discuss the creation of a k-nearest neighbor model and applications for the method.

## **Unit 5 >> Data Mining for Recommendation**

### **Introduction**

A data model is only as good as the data. Even with clean data, every data model has strengths and weaknesses. The business analyst must continue to improve data-mining models with experimentation and flexibility (changes in the objective) and then choose which model is better suited at a given point in time to meet the needs of the decision maker regarding the production of Business Intelligence.

Remember, a model is a representation of the "real world." In the real world, behaviors and tastes change daily. A data-mining model and its algorithms must change as new information becomes available and new Business Intelligence is needed. Changes include new business objectives, areas of focus, or patterns found in the data. The business analyst must be able to create and maintain data-mining models that can adapt to these changes and the model objectives, or else their recommendations to decision makers may be flawed.

## Learning Activities

### u05s1 - Studies

## Readings

Use the Capella University Library to read the following:

- In your *A Practical Guide to Data Mining for Business and Industry* text:
  - Part II, Section 4, "Data Preparation," pages 60–77.
  - Part III, Section 13.3, "Differences Between Statistical Analysis and Data Mining," pages 275–277.
- Bedeian, A. G. (2014). "More than meets the eye": A guide to interpreting the descriptive statistics and correlation matrices reported in management research. *Academy of Management Learning and Education*, 13(1), 121–135.
- Bălița, S., Žuka, R., & Krasts, J. (2016). Opportunities for the use of business data analysis technologies. *Economics and Business*, 28, 20–25.

## Videos

Complete the following brief instructional video.

- Miller, W. (2016). *Python for data science: Data preprocessing and text mining [Video]*. Skillsoft Ireland Limited.

### u05s1 - Learning Components

- Give examples of the different types of BI software applications used by business to help decision makers.
- Describe the different data modeling and statistical software applications available.
- Explain the issues that relate to descriptive business statistics and data mining processes within an organization.
- Explain the impact of legal and ethical issues of descriptive business statistics and data mining that can affect a company's ability to compete.
- Describe how data bias can affect the relevance and usability of descriptive business statistics and data mining.

### u05s2 - Optional: Software Download and Installation

You may download and use any of following software tools, as appropriate, for assignments in this course.

- SAS Enterprise Guide.
- SAS Enterprise Miner.
- Tableau Desktop.
- Qlik Sense Desktop.
- Analytic Solver (formerly XLMiner).

Directions for downloading the software are provided in the Unit 1 studies.

### Course Resources

---

Optional software: Analytic Solver, Qlik Sense Desktop, SAS Enterprise Guide, SAS Enterprise Miner, or Tableau Desktop. Directions for downloading the software are provided in the Unit 1 studies.

## u05a1 - Business Analytics to Support BI Strategy

By successfully completing this assignment, you will demonstrate your proficiency in the following course competencies and assignment criteria:

- Competency 1: Evaluate the process and workflows used in data interpretation and analysis.
  - Explain the issues that relate to descriptive business statistics and data mining processes within an organization.
- Competency 2: Evaluate the quality and constraints of source data used to solve business problems.
  - Explain the impact of legal and ethical issues of descriptive business statistics and data mining that can affect a company's ability to compete.
  - Describe how data bias can affect the relevance and usability of descriptive business statistics and data mining.
- Competency 4: Use statistical software to manipulate data to solve business problems.
  - Give examples of the different types of software applications used by business to help decision makers.
  - Describe the different data modeling and statistical software applications available.
- Competency 5: Communicate effectively.
  - Communicate effectively.

## Instructions

This assignment will focus on a U.S.-based business of your choosing. The business may be real or fictional, a start-up or well established, as long as it is of sufficient size to need data mining software for the type of business. The business you selected may focus on manufacturing, distribution, retailing, health care, service, or any combination of these purposes.

For this assignment, research and write a report as a newly hired business analyst. Assume you will submit your report to the director of analytics and strategy. Your work should be well documented and of use to both management and staff, particularly business intelligence managers and business analysts, as they set the tone for the careful setup and management of the company's data mining projects. You should draw on the concepts and processes you have learned in this course about Business Intelligence.

Your 6–10-page report, which should follow APA format, should include the following:

- Describe the core processes of the business and identify how these processes relate to descriptive business statistics and data mining. Be sure to address the quality of data stored.
- Use a program like Microsoft Visio to model the business flows related to the updated business analytics. Include the chart as an inline illustration or as an appendix to this paper. You may use SAS Enterprise Guide or Enterprise Miner.
- Evaluate the impact of descriptive business statistics and data mining on developing effective processes for decision makers.
- Identify the processes within an organization that involve the use of descriptive business statistics and data mining, along with the source (or sources) and plan for data integration.
- Explain the issues related to descriptive business statistics and data mining processes within an organization.
- Provide a detailed example of data mining in the context of the chosen business, as well as the range of activities that might fall under the umbrella of data mining.
- Research the critical success factors specific to particular data mining activities.
- Apply a specific data mining activity within context.
- Examine any legal and ethical issues related to specific examples of core data processes and how these issues affect descriptive business statistics and data mining. Illustrate how these issues may impact the business's competitiveness.
- Describe how data bias can affect the relevance and usability of descriptive business statistics and data mining.

Be sure to address the scoring guide criteria in your analysis.

## Other Requirements

- **Written communication:** Written communication should be free of errors that detract from the overall message.
- **APA formatting:** Your paper should be formatted according to the current APA style and formatting and should include a cover page and table of contents.
- **Length:** 6–10 typed and double-spaced pages, plus a cover page, table of contents, and appendix.
- **Font and font size:** Times New Roman, 12 point.

Course Resources

---

[APA Style and Format](#)

Provide examples of the different types of software applications that produce Business Intelligence used by businesses to help decision makers. How many of the different software applications provide decision-tree and k-nearest neighbor modules? What are some advantages and disadvantages of data mining applications? Which one produces more actionable Business Intelligence?

## Response Guidelines

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

Course Resources

---

Undergraduate Discussion Participation Scoring Guide

u05d1 - Learning Components

- Give examples of the different types of BI software applications used by business to help decision makers.
- Compare and contrast BI software applications.

**u05d2 - Course Reflection**

Reflect on the past five weeks. Share with the other learners what you have found most informative about the course and how you can apply what you have learned in future endeavors. Did you find the production of Business Intelligence regarding specific decision-making processes an easy task? What is your preferable analytic method, if any?

## Response Guidelines

Read your peers' initial discussion posts and respond to at least two of them. Compare your post to those of your peers and note any differences. Explain why you agree or disagree with your peers' views and analyses. Your responses are expected to be substantive in nature and should reference the assigned readings or other professional literature, as applicable, to support your views.

Course Resources

---

Undergraduate Discussion Participation Scoring Guide

u05d2 - Learning Components

- Reflect on and discuss the production of Business Intelligence and preferred analytic methods