

## **Syllabus**

### **Course Overview**

The objective of this course is to provide you with both the theory and practice of data analytics, thus enabling you to apply what you learn here in class to your current or future work. It is important to understand that the results obtained through data analytics is only as good as the data for which the analysis is based. Data mining is one of many methods used in data analytics to add value by adding new knowledge.

Traditional "statistics is the science of collecting, describing, and analyzing the data" (Lock, Lock, Lock Morgan, Lock, & Lock, 2013, p. 4). This is done by creating a mathematical model that fits and tests the data in a predefined way (Mirkin, 2011). Data mining, on the other hand, uses data to add new knowledge and insights by finding patterns that may be useful to the organization (2011).

As analysts, you may hear different terms when it comes to data mining, depending on who the data analyst is and which discipline they come from. The various terms you may hear are concepts, cases, objects, attributes, characters, states, and variables. Concepts, cases, and objects are synonymous with each other. Variables, attributes, characters, and states are synonymous with each other. For this course, we will use cases and 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 learn and group cases, whereas statistics tends to focus mostly on quantitative summarization. Both methods and variables are important to the data analyst.

A typical request of a data analyst is to understand the definition of a problem, review the problem, and then looks at the data and/or information that may be useful as the best answer to the problem. However, the job of the data analyst is often to look at some data and make sense of it in order for others to understand.

A data set comprises of two parts: data and meta data.

Meta data tells something about the data.

### References

Alexander, M., Decker, J., & Wehbe, B. (2014). *Microsoft business intelligence tools for Excel analysts*. Hoboken, NJ: John Wiley & Sons.

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.

## Course Competencies

(Read Only)

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

- 1 Define the role that data structure and distribution plays in data mining specific to an IT environment.
- 2 Apply data quality and cleansing tools to prepare data appropriate for a specific data mining project.
- 3 Apply spreadsheets and pivot tables to manipulate data for use in a data mining project.
- 4 Use SAS to support a specific mining project.
- 5 Communicate effectively.

## Course Prerequisites

IT2230.

**Syllabus >> Course Materials**

**Required**

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

**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.

- Alexander, M., Decker, J., & Wehbe, B. (2014). [Microsoft business intelligence tools for Excel analysts](#). Hoboken, NJ: John Wiley & Sons.
- Arora, J., Bhalla, N., & Rao, S. (2013). [A review on association rule mining algorithms](#). *International Journal of Innovative Research in Computer and Communication Engineering*, 1(5), 1246–1251.
- Azzalini, A., & Scarpa, B. (2012). [Data analysis and data mining: An introduction](#). Cary, NC: Oxford University Press.
- Balik, R. J. (2009). [Excel best practices](#). *Managerial Finance*, 35(5), 410–426.
- De Mulder, W. W. (2011). [Generalized hard cluster analysis](#). *International Journal of Computer Mathematics*, 88(12), 2516–2526.
- Emblemavåg, J. (2005). [Business analytics: Getting behind the numbers](#). *International Journal of Productivity and Performance Management*, 54(1), 47–58.
- Halfens, R. G., & Meijers, J. M. (2013). [Back to basics: an introduction to statistics](#). *Journal of Wound Care*, 22(5), 248–251.
- Hern, M. A., & Stolfo, S. J. (1998). [Real-world data is dirty: Data cleansing and the merge/purge problem](#). *Data Mining and Knowledge Discovery*, 2(1), 9–37.
- Hirji, K. K. (2001). [Exploring data mining implementation](#). *Communications of the ACM*, 44(7), 87–93.
- Mahafzah, B. A., Al-Badarneh, A. F., & Zakaria, M. Z. (2009). [A new sampling technique for association rule mining](#). *Journal of Information Science*, 35(3), 358–376.
- March, S. T., & Hevner, A. R. (2007). [Integrated decision support systems: A data warehousing perspective](#). *Decision Support Systems*, 43(3), 1031–1043.
- Mishra, V., Mishra, T. K., & Mishra, A. (2013). [Algorithms for association rule mining: A general survey on benefits and drawbacks of algorithms](#). *International Journal of Advanced Research in Computer Science*, 4(8), 155–159.
- O'Donnell, P. (2005). [The problem with pivot tables](#). *Business Intelligence Journal*, 10(1), 25–31.
- Rahman, N., Marz, J., & Akhter, S. (2012). [An ETL metadata model for data warehousing](#). *Journal of Computing & Information Technology*, 20(2), 95–111.
- Refaat, M. (2006). [Data preparation for data mining using SAS](#). Burlington, MA: Morgan Kaufmann.
- Saad, G. H. (2001). [Strategic performance evaluation: Descriptive and prescriptive analysis](#). *Industrial Management + Data Systems*, 101(8), 390–399.
- Saunders, J. (1994). [Cluster analysis](#). *Journal of Marketing Management*, 10(1–3), 13–28.
- Shu-Hsien, L., Pei-Hui Chu, Pei-Yuan Hsiao. (2012). [An empirical investigation of factors influencing the adoption of data mining tools](#). *International Journal of Information Management*, 32(3), 257–270.
- Tuffery, S. (2011). [Data mining and statistics for decision making](#). Hoboken, NJ: John Wiley & Sons.
- Tyagi, S. (2003). [Using data analytics for greater profits](#). *The Journal of Business Strategy*, 24(3), 12–14.
- Udiya, H., Kabra, R., & Easo, S. (2012). [An overview of association rule mining techniques](#). *Journal of Current Engineering Research*, 2(5), 62–66.
- Xiaohua, H., & Cercone, N. (2004). [A data warehouse/online analytic processing framework for web usage mining and business intelligence reporting](#). *International Journal of Intelligent Systems*, 19(7), 585–606.

## 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.

- Abousalh-Neto, N. (2013). [The forest and the trees: See it all with SAS Visual Analytics](http://support.sas.com/resources/papers/forest-trees.pdf) (SAS Institute Inc., Paper 058-2013). Retrieved from <http://support.sas.com/resources/papers/forest-trees.pdf>
- DelGobbo, V. (2008). [Tips and tricks for creating multi-sheet Microsoft Excel workbooks the easy way with SAS](http://www2.sas.com/proceedings/forum2008/192-2008.pdf) (SAS Institute Inc., Paper 192-2008). Retrieved from <http://www2.sas.com/proceedings/forum2008/192-2008.pdf>
- DelGobbo, V. (2009). *More tips and tricks for creating multi-sheet Microsoft Excel workbooks the easy way with SAS* (SAS Institute Inc., Paper 152-2009). Retrieved from <http://support.sas.com/resources/papers/proceedings09/152-2009.pdf>
- Gour, V., Sarangdevot, S.S., Tanwar, G.S., Sharma, A. (2010). [Improve performance of extract, transform and load \(ETL\) in data warehouse](http://www.enggjournals.com/ijcse/doc/IJCSE10-02-03-108.pdf). *International Journal on Computer Science and Engineering*, 2(3), 786–789. Retrieved from <http://www.enggjournals.com/ijcse/doc/IJCSE10-02-03-108.pdf>
- Lee, M. L., Ling, T. W., Lu, H. J., & Ko, Y. T. (1999). [Cleansing data for mining and warehousing](http://staff.icar.cnr.it/manco/Teaching/2006/datamining/articoli/dexa99.pdf). Retrieved from <http://staff.icar.cnr.it/manco/Teaching/2006/datamining/articoli/dexa99.pdf>
- Lee, T., Duling, D., Liu, S., & Latour, D. (2008). [Two-stage variable clustering for large data sets](http://support.sas.com/resources/papers/sgf2008/2stagecluster.pdf) (SAS Institute Inc., Paper 320-2008). Retrieved from <http://support.sas.com/resources/papers/sgf2008/2stagecluster.pdf>
- Rahm, E., & Do, H. (2000). [Data cleaning: Problems and current approaches](http://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf). Retrieved from <http://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf>

## 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.

## 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.

- Noonan, J. (2000). [Data mining strategies](http://www.information-management.com/issues/20000701/2367-1.html). *Information Management*. Retrieved from <http://www.information-management.com/issues/20000701/2367-1.html>

## Unit 1 >> Data Structure and Distribution

### Introduction

Data mining starts by agreeing and understanding the objective of the decision-makers. What exactly is the objective of the study? In many cases, the objective is to make sense of the data. What can the data tell decision-makers? What is the data story? Upon the agreement and understanding of the objective(s) by the data analyst, the next step is to gather and integrate the data. This step is often overlooked by many in both business and technology. Most of the time, the data is already available in data structures contained within source systems, operational data sources, data marts, or data warehouses. Even so, at a minimum, the data analyst must understand what the data is, how the data is used, and where the data comes from. Knowing this information about the data will help the data analyst with their exploratory analysis. For this unit, we are going to cover the basics of statistics used in for descriptive analysis and the influence data mining has on data analytics.

### Learning Activities

#### u01s1 - Studies

##### Readings

Use the following required texts found to complete the following:

Use your [Data Analysis and Data Mining](#) text to read the following:

- Chapter 1: "Introduction," pages 1–14.

Use your [Data Preparation for Data Mining Using SAS](#) text to read the following:

- Chapter 1: "Introduction," pages 1–5.

Use your [Data Mining and Statistics for Decision Making](#) text to read the following:

- Chapter 1: "Overview of data mining," pages 1–24.
- Chapter 3: "Data exploration and preparation," pages 43–91.

Use the Internet to read the following:

- Lee, M. L., Ling, T. W., Lu, H. J., & Ko, Y. T. (1999). [Cleansing data for mining and warehousing](http://staff.icar.cnr.it/manco/Teaching/2006/datamining/articoli/dexa99.pdf). Retrieved from <http://staff.icar.cnr.it/manco/Teaching/2006/datamining/articoli/dexa99.pdf>
- Rahm, E., & Do, H. (2000). [Data cleaning: Problems and current approaches](http://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf). Retrieved from <http://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf>

## Resources

### u01s2 - SAS E-Learning and Software

#### Accessing Your SAS E-Learning Materials and SAS Certification Preparation Courses

As a business intelligence or data analytics learner at Capella University, you have access to a comprehensive list of SAS e-learning materials and SAS certification preparation courses. Whether you are interested in becoming SAS certified or you are just looking for additional resources to help advance your understanding of a particular SAS tool, simply activate your Capella SAS account to gain access to resources that will help you advance in the field of data analytics.

- [How to Activate Your Capella SAS Account](#).
- [Your E-Learning Material and Certification Preparation Courses](#).

#### Access SAS Analytics U Community and Software Tools

Once you create your Capella SAS profile account, you become a member of the SAS Analytics U community. This is an online interactive community where you will find valuable resources and free software tools.

#### SAS University Edition

As a member of the SAS Analytics U, you can download your own free version of SAS University Edition or gain Web access to more robust and powerful SAS tools by registering for SAS OnDemand. Use this link to learn more about SAS University Edition and download your own copy of SAS University Edition.

- [SAS University Edition](#).

#### SAS OnDemand Tools

To obtain cloud access to more powerful and robust SAS Tools, you must first [register with SAS OnDemand](#).

Once you have registered for SAS OnDemand and have obtained your user ID, simply click the link below that corresponds with the tool that you would like to access.

- [SAS Enterprise Minor](#).
- [SAS Studio](#).
- [SAS Enterprise Guide](#).

- [SAS Forecast Studio](#). The name of the Environment is C867a0717de594199827634ad02887e5b.
- [SAS JMP](#).

As you work your way through this course and through your program, lean on the [SAS Analytics U Community](#) to connect with fellow SAS users. Reach out other community members to obtain peer support, get questions answered, share ideas and best practices, and collaborate on projects.

Another available resource is the [SAS OnDemand for Academics User's Guide](#).

## Resources

### u01a1 - Introduction to Data Structures' Impact on Data Mining

Most of the time, the data is already available in data structures contained within source systems, operational data sources, data marts, or data warehouses. Even so, at a minimum, the data analyst must understand what the data is, how the data is used, and where the data comes from. Knowing this information about the data will help the data analyst with their exploratory analysis.

For the Unit 1 assignment, you will research and write a short (3–4 pages for the body section) paper in APA (6th edition) style and format, with a minimum of five references, which covers the following topics:

1. Explain why it is important to understand the goals of a data mining study before starting to look at the data and data structures.
2. Identify and explain the role that data structure and distribution plays in data mining within an IT context.
3. Explain how appropriate data quality processes and best practices help data mining project.

## Assignment Requirements

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA (6th edition) style and formatting.
- **Length of paper:** 3–4 pages, excluding the references page.
- **Font and font size:** Times New Roman, 12 point.

## Resources

-  [APA Style and Format](#).

## Course Resources

[APA Style and Format](#)



## u01d1 - Purpose of Data Structures, Analytics, and Statistics

Using this unit's readings as a basis, discuss how data structures that contain the data help data analysts to understand and use the data for their data mining efforts. What do data analysts need to understand about the data structures? What are the differences and similarities between data analytics and classical statistics disciplines?

### Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

### Resources

#### Course Resources

[Undergraduate Discussion Participation Scoring Guide](#)

## u01d2 - Data and Data Structure Roles in Data Mining

Share an example, found in literature, where data structures can help or hinder a data analyst's ability to build, test, implement, and learn from their data mining models.

### Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

### Resources

#### Course Resources

[Undergraduate Discussion Participation Scoring Guide](#)

## Unit 2 >> Data Quality and ETL: Why It Matters

### Introduction

Extracting, transforming, and loading data from one system to a data mart or data warehouse or another database is only as good as the data definition and data rules that the process is built off. Good data quality is comprised of valid, right, and the correct representation of data at all times (Olson, 2003). Only when the data is valid, right, and correct does it meet the definition of accurate data. A solid ETL (extract, transform, load) process design and implementation should have the objective to handle 100% of the data that goes through the data pipeline for delivery to decision-makers. Too often, most ETL processes account for only the "happy path," due to a divide between functions and teams about who is responsible for data quality. Data quality can either positively or negatively affect the organization's data assets. For this unit, we are going to review data quality, the tools, and the processes used to create and manage a solid and reliable ETL process, while looking at some of the basic prep work of data that data mining studies use for testing and trials.

### References

Olson, J. (2003). *Data quality: The accuracy dimension*. San Francisco, CA: Morgan Kaufmann Publishers.

### Learning Activities

#### u02s1 - Studies

### Readings

Use the following required texts found within the ProQuest ebrary for Unit 2 studies:

Use your [Microsoft Business Intelligence Tools for Excel Analysts](#) text to read the following:

- Chapter 1: "Important Database Concepts," pages 1–18.

Use your [Data Preparation for Data Mining Using SAS](#) text to read the following:

- Chapter 2: "Tasks and Data Flow," pages 7–13.
- Chapter 6: "Integrity Checks," pages 63–82.

Use the Capella Library to read the following:

- Hern, M. A., & Stolfo, S. J. (1998). [Real-world data is dirty: Data cleansing and the merge/purge problem](#). *Data Mining and Knowledge Discovery*, 2(1), 9–37.
- March, S. T., & Hevner, A. R. (2007). [Integrated decision support systems: A data warehousing perspective](#). *Decision Support Systems*, 43(3), 1031–1043.

- Rahman, N., Marz, J., & Akhter, S. (2012). [An ETL metadata model for data warehousing](#). *Journal of Computing & Information Technology*, 20(2), 95–111.
- Shu-Hsien, L., Pei-Hui Chu, Pei-Yuan Hsiao. (2012). [An empirical investigation of factors influencing the adoption of data mining tools](#). *International Journal of Information Management*, 32(3), 257–270.
- Tyagi, S. (2003). [Using data analytics for greater profits](#). *The Journal of Business Strategy*, 24(3), 12–14.
- Xiaohua, H., & Cercone, N. (2004). [A data warehouse/online analytic processing framework for web usage mining and business intelligence reporting](#). *International Journal of Intelligent Systems*, 19(7), 585–606.

Use the Internet to read the following:

- Gour, V., Sarangdevot, S.S., Tanwar, G.S., Sharma, A. (2010). [Improve performance of extract, transform and load \(ETL\) in data warehouse](#). *International Journal on Computer Science and Engineering*, 2(3), 786–789. Retrieved from <http://www.enggjournals.com/ijcse/doc/IJCSE10-02-03-108.pdf>

### Optional Article

You can access the following article by registering at the *Information Management* Web site:

- Noonan, J. (2000). [Data mining strategies](#). *Information Management*. Retrieved from <http://www.information-management.com/issues/20000701/2367-1.html>

### Resources

## u02a1 - Data Quality & Cleansing Tools' Effect on Data Mining

Often, people think that ETL (extract, transformation, and load) is all there is to ensure data quality. There is a lot more to data quality than ETL; however, a data analyst should be familiar with ETL basics: processes, techniques, and tools. Data mining models may not perform well with inaccurate data or dirty data. The time to train and test a model may cause a project to fail when data is sparse; sparse data may lead to more time during exploration and finding better data to use for training. Understanding the basics of data management to include data quality may help a data analyst take less time to succeed with their data mining project.

For the Unit 2 assignment, you will research and write a short (3–4 pages for the body section) paper in APA (6th edition) style and format, with a minimum of five references, that covers the following topics:

1. Explain how extract, transform, and load (ETL) can affect data quality, data management goals, and affect data mining projects positively and negatively.
2. Describe how cleaning tools are used to prepare data for data mining projects.
3. Explain how to use SAS to create tasks and data flows.
4. Explain how to ensure data integrity can enhance a data model with SAS.

## Assignment Requirements

**Written communication:** Written communication is free of errors that detract from the overall message.

**APA formatting:** Resources and citations are formatted according to APA (6th edition) style and formatting.

**Length of paper:** 3–4 pages, excluding the references page.

**Font and font size:** Times New Roman, 12 point.

### Resources

-  [APA Style and Format](#).

#### Course Resources

[APA Style and Format](#)

## u02d1 - Data Mining: Dependence on Data Quality

Using this unit's readings as a basis, discuss how data mining modeling depends on the quality of data. What do data analysts need to understand about data quality? What are some ETL best practices that, when implemented correctly, can improve the outcome of a data mining project? How can data mining improve data quality and an ETL process?

### Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

### Resources

#### Course Resources

Undergraduate Discussion Participation Scoring Guide

## u02d2 - Data Integrity Improvement With Data Analytics Tools

Share an example, found in literature, where SAS or another data analytics tool was used to improve data quality. How was data integrity improved with the use of SAS or a data analytic tool?

### Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

### Resources

#### Course Resources

Undergraduate Discussion Participation Scoring Guide

## Unit 3 >> Manipulating Data: An Introduction With Excel Pivot tables

### Introduction

Pivot Tables, especially, Microsoft Excel PivotTables, are a powerful tool for data analysts. The ability for a data analyst to shift the view of a data set may allow for additional discoveries and patterns within the data itself. Pivot tables can help with the visual overview of a data set with both tables and charts. In this unit, we will cover the basics of pivot tables. You will learn how to create and use a pivot table using Microsoft Excel for analysis, which should help with making recommendations to decision-makers.

Experience with the following Excel basics will help you this unit:

- Importing and exporting data.
- Formatting data.
- Working with Excel functions.
- Filtering and sorting data.

There are links to the Excel resources on iGuide. If you need information on how to work with Excel, visit the [Microsoft Tutorials](#) page on iGuide. This page will take you to the Skillsoft portal, where you can access MS Excel tutorials.

### Learning Activities

### Readings

Use the following required texts found within the ProQuest ebrary for Unit 3 studies:

Use your [Microsoft Business Intelligence Tools for Excel Analysts](#) text to read the following:

- Chapter 2: "PivotTable Fundamentals." pages 19–46.

Use your [Data Preparation for Data Mining Using SAS](#) text to read the following:

- Chapter 7: "Exploratory Data Analysis," pages 83–97.

Use your [Data Mining and Statistics for Decision Making](#) text to read the following:

- Chapter 2: "The Development of a Data Mining Study," pages 25–42.

Use the Capella library to read the following:

- Balik, R. J. (2009). [Excel best practices](#). *Managerial Finance*, 35(5), 410–426.
- Emblemståg, J. (2005). [Business analytics: Getting behind the numbers](#). *International Journal of Productivity and Performance Management*, 54(1), 47–58.
- Halfens, R. G., & Meijers, J. M. (2013). [Back to basics: an introduction to statistics](#). *Journal of Wound Care*, 22(5), 248–251.
- Hirji, K. K. (2001). [Exploring data mining implementation](#). *Communications of the ACM*, 44(7), 87–93.
- O'Donnell, P. (2005). [The problem with pivot tables](#). *Business Intelligence Journal*, 10(1), 25–31.
- Saad, G. H. (2001). [Strategic performance evaluation: Descriptive and prescriptive analysis](#). *Industrial Management + Data Systems*, 101(8), 390–399.

### Resources

## u03a1 - Widgets for Which Customers

The board of directors wants additional information from analysis before making a final decision to move into a line of widgets for the upper Midwest region. Specifically, the board of directors wants to understand if a new line of widgets appeals to a certain age group of consumers. If so, the board of directors may allocate the company's valuable resources to produce, market, and distribute the new line of widgets to stores within the upper Midwest regions with the goal to obtain the greatest gross profit. The data within the **WidgetSpin.xlsx** contains the feedback received from a local study.

Your task is to analyze the study's results and report your findings with a recommendation to the board of directors using Excel and Excel Pivot Tables.

Assignment Requirements

Create an Excel Pivot Table with the data provided in **WidgetSpin.xlsx** and **WorkingWithExcelPivotTables.docx**. Add additional fields to the PivotTable. Customize the data as needed. Calculate the standard deviation and compare to the mean. Based off your analysis, what are your recommendations?

For the Unit 3 assignment, you will research and write a short (4–5 pages for the body section) paper in APA (6th edition) style and format, with a minimum of five references, that covers the following topics:

- 1. Explain the basic framework of a data analytics project.
- 2. Explain how bias can enter into spreadsheet models.
- 3. Demonstrate the basics of pivot tables.
- 4. Describe how central tendency statistics can help or hinder actions taken by decision-makers.
- 5. Share your recommendations based off the analysis of your Excel PivotTable work.

Resources

-  [APA Style and Format](#).

Course Resources
<a href="#">APA Style and Format</a>

u03d1 - Elements of Pivot Tables

Describe in your own words the basic elements of an analytics project strategy using pivot tables. What are important factors to consider when evaluating the results of a pivot table? Why is it important to understand the problem statement/objective before implementing a solution with pivot tables?

Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide

u03d2 - Pivot Tables Can Tell a Story

Give an example of an approach a business analyst can take to make sense of the data for decision-makers. How can pivot tables help with analysis to tell a data story to decision-makers? How does data mining help organizations?

Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide

Unit 4 >> Data Analytics and Mining With SAS

Introduction

Organizations that compete on analytics and reduce organizational politics and running their operations with "gut-feels" are pulling far ahead of their competitors (Davenport , 2006, 2005; Davenport & Harris, 2007, 2005). Organizations that are not competing on data analytics fall into two high-level groups. One group consists of those who want to compete on analytics and the other group is those who not want to compete on data analytics. Several studies break these groups further into different levels. So, what?

Data analysts are hired to help decision-makers understand their business environment. However, most of the time, data analysts spend more time cleaning and conforming the data than actually designing, testing, implementing, and learning from their data-mining models (Refaat, 2006). Tools like SAS and Microsoft Excel can help the data analyst find discrepancies and gaps within the source data. Once these issues are identified and addressed, the data analyst can create their data-mining model and verify its reliability with experimentation and additional testing.

References



- Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, 84(1), 98–107.
- Davenport, T. H. (2005). Analyze this: more and more companies are using analytics to drive their decision-making processes. But there's a right and a wrong way to do it. *CIO*, 19(1), 1–46.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Boston, MA: Harvard Business School Press.
- Davenport, T. H., & Harris, J. G. (2005). Automated decision making comes of age. *MIT Sloan Management Review*, 46(4), 83.
- Refaat, M. (2006). *Data preparation for data mining using SAS*. Burlington, MA: Morgan Kaufmann.

## Learning Activities

### u04s1 - Studies

#### Readings

Use your [Data Mining and Statistics for Decision Making](#) text to read the following:

- Chapter 6: "An Outline of Data Mining Methods," pages 167–174.

Use the Capella library to read the following:

- Arora, J., Bhalla, N., & Rao, S. (2013). [A review on association rule mining algorithms](#). *International Journal of Innovative Research in Computer and Communication Engineering*, 1(5), 1246–1251
- De Mulder, W. W. (2011). [Generalized hard cluster analysis](#). *International Journal of Computer Mathematics*, 88(12), 2516–2526.
- Mahafzah, B. A., Al-Badarnah, A. F., & Zakaria, M. Z. (2009). [A new sampling technique for association rule mining](#). *Journal of Information Science*, 35(3), 358–376.
- Mishra, V., Mishra, T. K., & Mishra, A. (2013). [Algorithms for association rule mining: A general survey on benefits and drawbacks of algorithms](#). *International Journal of Advanced Research in Computer Science*, 4(8), 155–159.
- Saunders, J. (1994). [Cluster analysis](#). *Journal of Marketing Management*, 10(1–3), 13–28.
- Udiya, H., Kabra, R., & Easo, S. (2012). [An overview of association rule mining techniques](#). *Journal of Current Engineering Research*, 2(5), 62–66.

Use the Internet to read the following:

- Abousalh-Neto, N. (2013). [The forest and the trees: See it all with SAS Visual Analytics \(SAS Institute Inc, Paper 058-2013\)](#). Retrieved from <http://support.sas.com/resources/papers/forest-trees.pdf>
- Lee, T., Duling, D., Liu, S., and Latour, D. (2008). [Two-stage variable clustering for large data sets \(SAS Institute Inc, Paper 320-2008\)](#). Retrieved from <http://support.sas.com/resources/papers/sgf2008/2stagecluster.pdf>

#### Resources

## u04a1 - Facts: What Facts Say About Products

Senior managements want the facts relating to results captured within a specific data set. What are the specific characteristics of the products contained in the data set? What is the trend of each product? What is the most likely outcome of these products in the near future?

### Assignment Requirements

Prepare and import the data from one of the "CARS" (SAS, 2014) Library files to create a new data source that you will use to run and evaluate your SAS analytics model. Use the data source that you just created and perform the following descriptive statistics: frequency, mean, and variant. Next, create a decision tree model using your SAS data source. Please open and follow the **UsingSAS\_for\_DescriptiveAndClusterAnalysis.doc** for directions on how to interact with SAS. Please answer the questions within the document highlighted in yellow and include the results of your analysis—screen prints and output reports—for review.

For the Unit 4 assignment, you will research and write a short (4–5 pages for the body section) paper in APA (6th edition) style and format, including screen prints, with a minimum of five references, that covers the following topics:

1. Explain the basic framework of a data analytics project using SAS.
2. Demonstrate the basics of using SAS to perform descriptive analytics.
3. Demonstrate the basics of using SAS to perform cluster analysis.
4. Explain the results of your SAS work using descriptive statistics and a decision tree model using SAS.
5. Share your recommendations based off your SAS analytics work.

### References

SAS. (2014). SAS Studio.

### Resources

-  [APA Style and Format](#).

### Course Resources

[APA Style and Format](#)

## u04d1 - Data Mining Methods Available with SAS

Describe how data quality can help or hinder clustering and association algorithms used by data analysts to help decision-makers. How does SAS transformation work to help data analysts create and use a data model? What data mining methods are available with SAS? How does understanding the basics of association or cluster algorithms help a data analyst?

### Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

### Resources

#### Course Resources

Undergraduate Discussion Participation Scoring Guide

## Unit 5 >> Data Analytics and Mining: Putting It All Together

### Introduction

The data model is only as good as the data. Even with clean and conformed data, data-mining models will have their strengths and weaknesses. The data analyst must continue to improve their data mining models with experimentation, flexibility (changes in the objective), and choose which model is better suited at a given point in time to meet the needs of the decision-maker.

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

Often, missing data (sparse), incorrect data, problem definitions, and the hopes or current understanding of decision-makers influences a data-mining project. Experience and knowledge about how to create and analyze various associations and clusters can help the data analyst provide facts and recommendations to decision-makers that could help them make better decisions for the organization.

### Learning Activities

### Readings

Use your [Data Preparation for Data Mining Using SAS](#) text to read the following:

- Chapter 10: "Binning and Reduction of Cardinality," pages 141–170.
- Chapter 11: "Treating of Missing Values," pages 171–204.

Use the Internet to read the following:

- DelGobbo, V. (2008). [Tips and tricks for creating multi-sheet Microsoft Excel workbooks the easy way with SAS](http://www2.sas.com/proceedings/forum2008/192-2008.pdf) (SAS Institute Inc, Paper 192-2008). Retrieved from <http://www2.sas.com/proceedings/forum2008/192-2008.pdf>
- DelGobbo, V. (2008). [More tips and tricks for creating multi-sheet Microsoft Excel workbooks the easy way with SAS](http://support.sas.com/resources/papers/proceedings09/152-2009.pdf) (SAS Institute Inc, Paper 152-2009). Retrieved from <http://support.sas.com/resources/papers/proceedings09/152-2009.pdf>

### Resources

## u05a1 - Competing on Data Analytics and Data Mining

Use the following scenario to complete your assignment.

Choose a business based in the United States. The business may be fictional or based upon an actual business—but it must be of sufficient size to need data mining software for the type of business. It may be a startup or existing company but in need of data analytics to help decision-makers understand what is going on within the business. The business may have as its mission manufacturing, distribution or retailing, health care, purely service, or any combination thereof.

You will be researching and writing this paper as the newly hired data analyst, reporting to the director of analytics and strategy of the company. Your goal is to write a documented paper that can be used both by management and staff, particularly data analytic and strategy managers and business analysts, to set the tone for the careful setup and management of the company's data mining projects. You will be using all concepts and processes you have learned in this course.

### Assignment Requirements

For this assignment, research and write a scholarly paper in APA format (approximate body length of 6–10 pages), in which you include the following.

1. Describe the core processes of the company and identify how those processes are related to descriptive statistics and data mining.
2. Use a program like Microsoft Visio to model the data flow(s) related to the updated data analytics. Include the chart as an inline illustration or as an appendix to this paper.
3. Evaluate the impact of descriptive statistics and data mining in developing effective processes for decision-makers.
  - Identify the processes within an organization that involve the use of descriptive statistics and data mining.
  - Explain the issues that relate to descriptive statistics and data mining processes within an organization.
  - Demonstrate how using Pivot Tables and SAS can help articulate what happened and what is going-on to help decision-makers understand their environment.
4. Illustrate an example of data mining.
  - Identify the range of activities that might fall under the umbrella of data mining.
  - Research the critical success factors specific to a particular data mining activities.
  - Apply a specific data mining activity within context.
5. Apply legal and ethical issues to specific examples of core data processes.
  - Identify legal and ethical issues that affect descriptive statistics and data mining.
  - Describe how data bias can affect the relevance and usability of descriptive statistics and data mining.
  - Illustrate the impact of legal and ethical issues of descriptive business statistics and data mining that can affect a company's ability to compete on data analytics.

#### Resources

-  [APA Style and Format](#).

#### Course Resources

[APA Style and Format](#)

### u05d1 - Reducing the Noise With SAS

Describe how to handle sparse data (missing values) within a data-mining project using SAS. How can a data analyst reduce the noise caused by sparse data? What are some causes of sparse data within a data-mining project? How can transformation of source data help or hinder a data-mining project? Explain how Microsoft Excel and SAS can be used together to help to reduce the noise caused by missing values.

Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide

u05d2 - 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.

Response Guideline

Read the posts of your peers and respond to two (minimum) and expand on the concepts covered in their initial post. Quantity and quality of your posts will determine the value of the group's learning experience. Provide a substantive and appropriate response.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide