

Syllabus

Course Overview

In this course, you will be working on two end-to-end advanced analytics solutions: one of your own choosing and one for Vila Health, as part of the virtual internship and related group work. The data will be provided for the Vila Health analysis, but for the individual hands-on assignments, you will need to have your own data already, or find or collect data to work with. The choice of topic is completely up to you, but you have to be able to gather, find, or use existing data (either at your place of employment, online, or via a survey, for example).

The requirements for your data set are:

1. It has to be usable to solve a business, personal, or other problem using predictive analytics.
2. It has to be raw data that is appropriate to use for predictions (transaction-level, account-level, day-level, for example).
3. It has to have at least one response variable that you are going to try to predict either explicitly or via a transformation or calculation(s).
4. It has to include at least 20 different options for explanatory variables (the exercises in this course include selecting a subset of variables, so you will need enough to be able to select a subset from).

Course Competencies

(Read Only)

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

- 1 Explain visual analytic approaches.
- 2 Explain geospatial data techniques.
- 3 Apply analytic skills to current organizational problems.
- 4 Apply analytic solution scoring.
- 5 Apply project management skills for effective team performance.
- 6 Present the results to stakeholders in a succinct and relevant manner.

Course Prerequisites

Prerequisite(s): ANLT5050.

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.

- Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons.
- Goben, A., & Raszewski, R. (2015). [The data life cycle applied to our own data](#). *Journal of the Medical Library Association*, 103(1), 40–44.
- Kiryluk, K., Li, Y., Sanna-Cherchi, S., Rohanizadegan, M., Suzuki, H., Eitner, F., . . . Gharavi, A. G. (2012). [Geographic differences in genetic susceptibility to IgA nephropathy: GWAS replication study and geospatial risk analysis](#). *PLoS Genetics*, 8(6), e1002765.

- Lin, L. (2010). *An ecological study of children commuting to school* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. 3443189)
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). *A survey of data mining and knowledge discovery process models and methodologies*. *The Knowledge Engineering Review*, 25(2), 137–166.
- Sharma, S., Osei-Bryson, K.-M., & Kasper, G. M. (2012) *Evaluation of an integrated knowledge discovery and data mining process model*. *Expert Systems With Applications*, 39(13), 11335–11348.
- Van Poucke, S., Zhang, Z., Schmitz, M., Vukicevic, M., Vander Laenen, M., Celi, L. A., & De Deyne, C. (2016). *Scalable predictive analysis in critically ill patients using a visual open data analysis platform*. *PLoS One*, 11(1), 1–22.

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.

- Belbin Worldwide. (n.d.). *Belbin team roles*. Retrieved from <http://www.belbin.com/rte.asp>
- LaLonde, S. M. (2012). *Transforming variables for normality and linearity— when, how, why and why not's*. Retrieved from <http://support.sas.com/resources/papers/proceedings12/430-2012.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.

- Kaggle. (n.d.). *Kaggle datasets*. Retrieved from <https://www.kaggle.com/datasets>
- *KD Nuggets*. (n.d.). Retrieved from <http://www.kdnuggets.com/datasets/index.html>
- *Python*. (n.d.). Retrieved from <https://www.python.org/>
- *R: The R Project for Statistical Computing*. (n.d.). Retrieved from <https://www.r-project.org/>

Unit 1 >> Overview of Advanced Analytics and Modeling

Introduction

In this unit, you will read several resources regarding getting started with an advanced analytics project, various data mining frameworks, and analytics lifecycles. In your group studies for this unit, you will select the roles you will play within your groups, set up communication channels, start to discuss the articles related to your virtual internship, and begin looking at the data that is available to you within Vila Health. This unit will serve as the basis for the rest of your individual work done during this course and will be where you provide the overview of the business, personal, or other problems you are attempting to solve using advanced analytics and modeling, as well as the details of the data you will be using to do so.

Remember, you will have two projects to work on throughout this course, the Data Analytics Internship: Vila Health project and the individual work project.

Learning Activities

u01s1 - Studies

Readings

From the Capella library, read:

- In Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons:
 - Chapter 1: Overview of Predictive Analytics.
 - Chapter 2: Setting Up the Problem.
- Goben, A., & Raszewski, R. (2015). *The data life cycle applied to our own data*. *Journal of the Medical Library Association*, 103(1), 40–44. This article provides an overview of a data analytical lifecycle from two library scientists' perspectives.
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). *A survey of data mining and knowledge discovery process models and methodologies*. *The Knowledge Engineering Review*, 25(2), 137–166.
- Van Poucke, S., Zhang, Z., Schmitz, M., Vukicevic, M., Vander Laenen, M., Celi, L. A., & De Deyne, C. (2016). *Scalable predictive analysis in critically ill patients using a visual open data analysis platform*. *PLoS One*, 11(1), e0145791.
- Sharma, S., Osei-Bryson, K.-M., & Kasper, G. M. (2012). *Evaluation of an integrated knowledge discovery and data mining process model*. *Expert Systems With Applications*, 39(13), 11335–11348.

Optional Internet Data Resources

These Web sites contain data you may find helpful in your individual project:

- Kaggle. (n.d.). [Kaggle datasets](https://www.kaggle.com/datasets). Retrieved from <https://www.kaggle.com/datasets>
- [KD Nuggets](http://www.kdnuggets.com/datasets/index.html). (n.d.). Retrieved from <http://www.kdnuggets.com/datasets/index.html>

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.

SAS Statistical Software

[SAS OnDemand for Academics \(SODA\)](#).

Open Source Statistical Software

R and Python are two open source software that can be applied as an alternative to SAS to complete the assignments in this course.

- **R**: Go to the [Download](#) page of the Getting Started section of The R Project for Statistical Computing home page to download the latest version of R.
- **Python**: Go to the [Download](#) section of the [Python Beginners Guide](#) to download the latest version of Python.

Selecting a Statistical Software

On the one hand, R and Python are open source software that are free and commonly used. On the other hand, there are several reasons why an analyst would choose SAS over R or Python. SAS has better reference documentation. R and Python are getting better, but official documentation is not well-written and somewhat opaque. SAS is supported by a massive infrastructure of books. SAS has single-source support. If it is broken, you know who to call. If R or Python are broken, there is nobody to call. R and Python packages vary in quality, some are written by experts, and some are not.

Whereas, SAS is only written by experts. SAS is very much the standard for many industries, especially the pharmaceutical industry, where SAS is required by the FDA. The name SAS carries a lot of inertia and is used predominantly in institutions and academia.

u01s3 - Programming Instructions

Using R

Access the [Using R](#) page on Campus for resources on how to get started with R.

Using Python

Access the [Using Python](#) page on Campus for resources on how to get started with Python.

Using SAS

Access the [Using SAS](#) page on Campus for resources on how to get started with SAS.

u01s4 - Analytics Internship: Advanced Analytics and Modeling

An integral part of this course and Capella University's data analytics program is your virtual internship. Your internship is with Vila Health, a fictional health care system that operates hospitals and other health care facilities throughout the upper Midwestern United States. This internship is designed to allow you to apply the skills and knowledge you are acquiring in a realistic setting. In addition to the technical requirements of the assignment, the internship also provides a context for you to develop the collaborative, interpersonal skills that employers are looking for in new data analysts.

In this first activity, you will be introduced to the Vila Health organization, your mentor, and the case study that you focus on throughout this course. You will use the information about the case study as the context for your group discussion Group Study Tasks in the coming weeks. Click **Analytics Internship: Advanced Analytics and Data Modeling** to begin.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling](#)

u01s5 - Group Work and Personal Effectiveness

This course adheres to Belbin's Team Theory and Belbin's Nine Team Roles. Review the resources provided on Belbin and collaboration, if needed. Later in this unit you will be asked to choose a role for the group work based on these theories.

You will work with your group in Unit 2 to choose a role for each group member. To prepare for this, decide now which two to three roles you feel you are best able to fill; preferably ones that you have not yet played in previous courses.

Study Group Tasks

For each week in the course, you will find suggested study group tasks describing topics to discuss with your group within a social media setting. These topics will help guide your group toward its midpoint and final solutions; therefore, you must not skip any of these study groups if you plan to succeed on the group project.

Review the most recent Vila Health challenge and any of the materials you created or acquired during project-related or challenge activities prior to your study group meetings. Downloaded notes from the notebook can provide useful reminders and questions for your group discussion. Be sure to consider the Belbin team role you are filling for your group and consider the information in the scenario from the perspective of that role or roles.

You will encounter study group tasks each week throughout the course, and in every course that you encounter as you journey through your program.

The questions provided are meant to provide you and your group members with some guidance for your group work deliverables. These questions should not be overlooked, as they are critical to you completing the work in the course.

Study Group Meetings

Capella recommends that you use a social media channel to collaborate and connect with your group members. Work with your teammates to identify a platform you are all comfortable using (for example, Basecamp, Yammer, Google Hangouts, or Facebook) and check to make sure that this platform is also reasonably private and secure.

For example, if you decide to use Facebook, create a secret group and work within it. A secret group is invisible to the outside world, and only members can find the group and see posts. This social media group will allow you to stay connected to your team and allow you to discuss the study group topics assigned.

LinkedIn

For networking with current and past learners, we recommend that you use LinkedIn, which is a great way to stay connected to past, current, and future learners in this degree program. This will help you with networking as you move through the program and beyond.

Course Resources

[Group Collaboration Skills](#)

Belbin Worldwide. (n.d.). [Belbin team roles](http://www.belbin.com/rte.asp). Retrieved from <http://www.belbin.com/rte.asp>

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

u01a1 - Problem Definition and Data Overview

If you have not done so already, please read the Course Summary for an overview of the two projects you will be completing in this course. This assignment pertains to your individual project.

For this assignment:

1. Write a 1–2 page summary of your selected industry/topic area, the business, personal, or other problem that you are trying to solve or question you are trying to answer. In this summary:
 - Explain why you selected this problem.
 - Include the source of the data that you are planning on using to address the problem or answer the question. If you are using a scenario from your place of work, remember to abide by any confidentiality agreements and get any prior approvals you need in order to utilize that data source. You will not need to include the actual data, and no PII data or any other confidential information will need to be shared for any of the assignments. General descriptions of variables, procedures, et cetera, should be sufficient, but remember that these assignments are not intended to be theoretical, so be sure that you are actually able to use and analyze the data you are describing.
2. Summarize your data in a table, which does not count as part of the 1–2 page summary, including:
 - A description of what a single record represents, for example, a single record represents biometric data from a single hour of one day; a single record represents one individual patient claim identified by patient ID and date; or a single record represents a single credit card transaction.
 - The name, description, and data type for each response variable that you are looking to predict.
 - The name, description, and data type for each variable that will be used as a candidate for an explanatory variable in your predictive model.

Your assignment will be scored on the following criteria:

- Explain a problem or question to be addressed.
- Justify the selection of problem or question and the importance of the analytics project for addressing this problem.
- Describe data available for solving a problem.

Refer to the Problem Definition and Data Overview Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 1–2 typed double-spaced pages, plus the data summary table.
- **Font and font size:** Arial, 10 point.

Course Resources

[APA Style and Format](#)

u01d1 - Planning Your Group

Throughout the course, you will be required to work collaboratively outside of the courseroom. By now, your instructor should have assigned you into groups of 2 to 3 members. With your group, be sure to:

- Introduce yourself to your group members. Share information about your background and experience and include your experience using and working with analytic software.
- Decide on the social media channels and/or collaborative tools that you as a group plan to use throughout this course.
- Meet with your group at your agreed-upon social media location and discuss this week's topics for discussion, as outlined in the group work study.
- Choose a Belbin Team Role for each member. Try to choose a different role for each member of the group.
- Review the Vila Health media piece and consider the scenario presented. Start considering the Vila Health scenario presented from the perspective of the Belbin team role or roles you are filling for your group.

Complete your group set up and role selection by Wednesday of this week.

Course Resources

[Graduate Discussion Participation Scoring Guide](#)

[Group Collaboration Skills](#)

[Belbin Team Roles](#)

Unit 2 >> Data Visualization Methods and Pattern Recognition

Introduction

Prior to initiating the building of the models themselves, it is imperative to ensure that we fully understand the contents of our data. We can do this through many summary processes and procedures, but we can also use visual methods to quickly identify outliers, check assumptions of distributions, and check for patterns in our data. In this unit, you will work with your group to continue to audit the data available for the Vila Health problem, and also audit the data for your individual project data set.

Learning Activities

u02s1 - Studies

Readings

From the Capella library, read:

- In Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons:
 - Chapter 3: Data Understanding.

u02s2 - Programming Instructions

Using R

Access the [Using R](#) page on Campus for resources on:

- How to get started with Descriptive Statistics in R.
- Getting started with Graphical Methods in R.
- Getting started with Model Building in R.

Using Python

Access the [Using Python](#) page on Campus for resources on:

- How to get started with Descriptive Statistics in Python.
- Getting started with Graphical Methods in Python.
- Getting started with Model Building in Python.

Using SAS

Access the [Using SAS](#) page on Campus for resources on:

- How to get started with Descriptive Statistics in SAS.
- Getting started with Graphical Methods in SAS.
- Getting started with Model Building in SAS.

u02s3 - Study Group Tasks

By this time, your group should have chosen an area in which to meet outside of the courseroom to discuss topics and formulate recommendations.

The recommendations that your group agrees to this week need to be included as part of the Midpoint Review and Final Report and will contribute to your final group project grade. Therefore, it is critical that you do not skip any of these discussions. You will be able to make changes to your recommendations at any time during the course, but skipping a topic will negatively impact your grade.

Review the Analytics Internship: Advanced Analytics and Data Modeling scenario. Consider the topics for discussion as you review the scenario. Search for details that are specific to these topics and that will aid you in making recommendations to Vila Health.

This week's topics for discussion:

- What variables will be used as response variables?
- What variables are potential options for use as explanatory variables in your model?
- What variables have missing data that you will have to deal with prior to creating your model?
- What other data quality issues (or potential data quality issues) did you identify in your initial evaluation of the data?
- What visualization methods might be appropriate for evaluating the values of each variable?
- What visualization methods might be appropriate for evaluating the relationships between the variables?
- What visualization methods might be appropriate for helping to identify related observations?

Keep in mind the role you are playing within your group. You may find that some information is more useful to your role than others.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

u02a1 - Data Audit Report

Overview

Once you have clearly defined the business problem you are attempting to address with your advanced analytics and modeling project and inventoried the data and variables available to you, it is time to dig in deeper and truly understand the data you are dealing with. In order to do this, you can use basic descriptive (summary) statistics, like measures of central tendency, measures of spread, measures of position for quantitative variables, and frequency or relative frequency tables for qualitative variables. You can also take this one step further to uncover aspects of the data and variables that might not be obvious with summary statistics and begin to visualize the data with various graphical methods.

Instructions

To complete this assignment:

1. Using the business problem and data that you inventoried in Unit 1, identify the variables that are candidates for response and explanatory variables in your model.
2. Use SAS to create a descriptive and visual analysis of each variable and to create a data audit report for your data. You will submit this report as the paper for this assignment. You should address the following for each variable within your data audit report:
 - What variables have missing values? In each of these variables with missing values, what percent of the values are missing?
 - For each quantitative variable:
 - Determine whether there any extreme values, or outliers, in each of the variables.
 - Describe the shape of the distribution of each variable. (How many modes does the distribution have? Where is the middle? How spread out are the variable values? Is there noticeable skewness to the distribution?)
 - For each qualitative variable:
 - Describe the cardinality of the qualitative variables (that is, how many unique values exist, or are possible values, in the qualitative variable).
 - Explain whether there are any strange or unexpected values in the qualitative variable.
 - Describe the distribution of values within the qualitative variable. Are there any values that show up way more or way less than others? Any other notable aspects about the distribution of values?
 - For relationships between variables:
 - Do any of the variables greatly impact or change the values of the response variable?
 - Do any of the variables have a strong relationship or correlation between each other?
3. Write a 4–6 page data audit report that summarizes your findings of your data audit and addresses the questions above. Writing the report in a table format usually makes the report easier to read.
4. Cite in-text if applicable and include references for any paraphrasing and direct quotations as well as any images that you use from the company, or other, sites or documentation. Not only does this provide credit to those sources, but, more importantly, it provides your reader with a resource with which to find additional information on the particular functionality or software.

Your assignment will be scored on the following criteria:

- Evaluate the quality of the data and variables available.
- Use visual methods to identify patterns in data.
- Identify response variables for a specific business problem.
- Identify explanatory variable candidates for a specific business problem.
- Use statistical software to create a descriptive and visual analysis of variables and to create a data audit report.

Refer to the Data Audit Report Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

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

- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 4–6 typed double-spaced pages, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

APA Style and Format

u02d1 - Visualization Methods

Using this unit's readings, compare and contrast the various visualization methods for evaluating data and variables prior to modeling, and answer the following questions:

1. What visualization methods are best for understanding and evaluating one variable?
2. What do these visualization methods provide over simple summary statistics?
3. What visualization methods are best for evaluating the relationship between two or more variables?
4. How do these visualization methods compare to correlation statistics or metrics, like the Pearson correlation coefficient, for example?
5. How might we use one or more visualization methods to understand whether we are dealing with multiple groups of observations within our data set?

Response Guidelines

Respond to at least two other learners and share with them the portions of their initial post that you agree or disagree with, and why.

Course Resources

Graduate Discussion Participation Scoring Guide

Unit 3 >> Data Preparation and Missing Values

Introduction

In our data audit, we evaluated all of the variables available to us, and identified potential issues with the data in these variables. In order to prepare our data for modeling, we need to address as many of the issues identified in the data audit as possible, even if addressing the issue only consists of identifying it as an issue and leaving the data as is. There are several options for dealing with missing values, including leaving them alone, making them 0 values, or imputing missing values that is, replacing the missing values with an estimated value of one type or another. There are also various reasons that one may want, or need, to transform a variable. These reasons may include the need to meet certain assumptions of the chosen modeling technique, like normality or linearity, for example.

In this unit, you will collaborate with your classmates and share ideas on dealing with missing values and discuss when transformations may or may not be needed on variables.

Learning Activities

u03s1 - Studies

Readings

From the Capella library, read:

- In Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons:
 - Chapter 4: Data Preparation.

From the Internet, read:

- LaLonde, S. M. (2012). *Transforming variables for normality and linearity—when, how, why and why not's*. Retrieved from <http://support.sas.com/resources/papers/proceedings12/430-2012.pdf>

u03s2 - Study Group Tasks

The recommendations that your group agrees to this week need to be included as part of the Midpoint Review and Final Report and will contribute to your final group project grade. Therefore, it is critical that you do not skip any of these discussions. You will be able to make changes to your recommendations at any time during the course, but skipping a topic will negatively impact your grade.

Review the Analytics Internship: Advanced Analytics and Data Modeling scenario. Consider the topics for discussion as you review the scenario. Search for details that are specific to these topics and that will aid you in making recommendations to Vila Health.

This week's topics for discussion:

- Are there missing values that we need to address? If so, what are the options for addressing them? What are the pros and cons of each option?
- Are the assumptions of our potential modeling techniques met? Are there normality assumptions or linearity assumptions that are not met? Are there any other assumptions that are not met?
- How do we plan to address any assumptions that are not currently met? Can these be addressed using a transformation? If so, what transformation is needed and on what variables?

Begin discussing the modeling techniques that may be appropriate for the Vila Health scenario and data, and identify the assumptions of those techniques. Use the data audit that your group conducted in Units 1 and 2 to evaluate whether or not the necessary assumptions are met, and, if not, begin discussing options for addressing these unmet assumptions. In addition, use the data audit to identify missing data that needs to be dealt with and begin planning how to address any missing data that your group identifies as needing to be addressed.

Keep in mind the role you are playing within your group. You may find that some information is more useful to your role than others.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

u03a1 - Data Cleansing and Modifications Specification

Overview

No universal solution exists for dealing with missing data, or for addressing unmet assumptions, that works for every situation and every set of data. For this reason, each option needs to be evaluated in each situation, and the optimal solution selected based on the business problem, the data available, and any other relevant factors. In this assignment, you will examine methods for dealing with missing values because we cannot predict a response variable based on a nonexistent value for an explanatory variable.

Instructions

To complete this assignment:

- Compare and contrast methods for dealing with missing values, including when each method is most and least appropriate.
- Choose and describe a method, or methods, for addressing any missing values in the data you are using for addressing your selected business problem and use SAS to implement your chosen methods.
- Use your data audit from the previous units to assess whether or not the assumptions are met for the modeling techniques you may be using to address your selected problem, and identify each unmet assumption.
- Evaluate options for addressing each unmet assumption, including, but not limited to, variable transformations or alternative modeling techniques, and select, and describe, one method for addressing each unmet assumption.
- Use SAS to implement your chosen method for addressing each unmet assumption.
- Write a 4–6 page paper that addresses each of the points above, and include the SAS code written and used in this assignment as an attachment or appendix (not included in the 4–6 pages).

Your assignment will be scored on the following criteria:

- Compare and contrast methods for dealing with missing data.
- Select a method, or methods, for dealing with missing data that is appropriate for a given problem.
- Assess whether assumptions for a modeling technique are met.
- Evaluate options for addressing unmet model assumptions.

Refer to the Data Cleansing and Modifications Specification Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 4–6 typed double-spaced pages, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

[APA Style and Format](#)

u03d1 - Variable Transformations and Addressing Missing Values

Use this discussion to collaborate on your individual projects by providing a summary or overview of any missing value problems you may have in your data, or by including transformations you think may be needed on one or more of your variables.

In your initial response, address the following:

1. Missing value problems:
 - What variables have missing values?
 - What percentage of the values are missing in each variable?
 - How do you propose to deal with these missing values?
2. Variable transformations:
 - Are there any variables in your data that are not distributed according to the assumptions of your modeling method? How do you propose to deal with the unmet assumptions?
 - What variables do you think might require transformations? Why?

Response Guidelines

Respond to at least two other learners and provide your feedback on their missing value issues and potential variable transformations needed. What do you agree with and what do you disagree with in their current approach, and why? If you disagreed with any portion of their approach, what other recommendations do you have? Remember to support your recommendations with references, so your classmates can conduct further research regarding your recommendations to determine if that approach will work for their unique situation.

Course Resources

Unit 4 >> Association Rules

Introduction

Sometimes we find situations where certain items are commonly purchased together or otherwise associated in some way. Other times, medications need to be taken with other medications or vitamins to maximize their effects. As another example, we may have a doctor with a particular specialty—a surgeon, for example—who will always be found with specific doctors that he needs to assist him, like an anesthesiologist.

Any time we have items that commonly go together in modeling, we call them *item sets*. Not every advanced analytics and modeling scenario lends itself to the use of item sets, but it is an important concept for us, as analysts, to understand, because it impacts the independence of variables related to those items and adds other considerations to our analysis related to the interaction between these items. Once we have identified the fact that a certain set of items commonly comes together, and have identified the appropriate item sets, we can take this one step further to create rules that reflect how these sets of items are related. These are often referred to as *association rules*.

In this unit, you will explore the concepts of item sets and association rules, read about applications of them, determine whether they are appropriate for your individual analysis and modeling project, and discuss with your group whether there may be item sets to consider in your Vila Health project.

Learning Activities

u04s1 - Studies

Readings

From the Capella library, read:

- In Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons:
 - Chapter 5: Item Sets and Association Rules.

u04s2 - Analytics Internship: Applied Forecasting



[Analytics Internship: Applied Forecasting](#)

 [Transcript](#)

In this Vila Health activity, you will continue working on the business problem introduced in earlier in the course. Click **Analytics Internship: Applied Forecasting** to view the Vila Health scenario.

Course Resources

[Analytics Internship: Applied Forecasting](#)

u04s3 - Study Group Tasks

The recommendations that your group agrees to this week need to be included as part of the Midpoint Review and Final Report and will contribute to your final group project grade. Therefore, it is critical that you do not skip any of these discussions. You will be able to make changes to your recommendations at any time during the course, but skipping a topic will negatively impact your grade.

Review the Analytics Internship: Applied Forecasting scenario. Consider the topics for discussion as you review the scenario. Search for details that are specific to these topics and that will aid you in making recommendations to Vila Health.

This week's topics for discussion:

- Are there any applications for item sets and association rules in the Vila Health scenario? If so, where and what types?
- If applicable, how do we go about finding the item sets and developing the related association rules in the Vila Health scenario?

Keep in mind the role you are playing within your group. You may find that some information is more useful to your role than others.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

u04a1 - Initial Individual Project Report

Overview

As an analyst, especially an entry-level or mid-career analyst, because of the length of effort involved in analytics and modeling projects, you are often required to plan for code reviews, intermediate reports, and other checkpoints with your manager and teammates in order to ensure that you are on the right track with everything.

This assignment is a way for you to submit to your instructor what you have done on your individual project so far. It should include all your completed work for your individual analytics and modeling projects in Units 1 through 4, and support all of your decisions and recommendations with appropriate references.

Instructions

For this assignment, write a paper to summarize the work you completed on your individual project to this point. This summary should include the following:

- Define the problem and the source of the data and explain how they align or do not align and what you will do to rectify any misalignment.
- Audit and profile the data and identify any potential data issues. Include any outputs, graphs, et cetera, from your data profiling and audit efforts.
- Address the data issues, including missing values and erroneous data. Include a specification and code for all data manipulation you had to complete in order to address these data issues.
- Write a 4–6 page paper that addresses each of the points above, and include the SAS code written and used in this assignment as an attachment or appendix (not included in the 4–6 pages).
- Support your methods and choices with references, where appropriate.

Your assignment will be scored on the following criteria:

- Define a problem and the source of the data and explain how they align or do not align with what needs to be done to rectify misalignment.
- Audit and profile the data and identify any potential data issues and include relevant outputs and graphs from data profiling and audit efforts.
- Address the data issues, including missing values and erroneous data and include a specification and code for all data manipulation completed in order to address data issues.
- Support selected data analytics methods and choices with references, where appropriate.

Refer to the Initial Individual Project Report Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 4–6 typed double-spaced pages, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

APA Style and Format

u04d1 - Association Rules and Their Applications

From this unit's reading and from research in the Capella library, define in layman's terms association rules and item sets, and provide at least one example of an application of same from an industry of your choice.

Response Guidelines

Respond to at least two other learners and share with them the portions of their initial post that you agree or disagree with, and why.

Course Resources

Graduate Discussion Participation Scoring Guide

Unit 5 >> Grouping Variables and Observations

Introduction

To this point, you have reviewed the concept of advanced analytics and modeling, and, for both your individual project and your Vila Health group project, have worked on defining the business problem, identifying the necessary data, auditing and profiling that data to identify issues, and cleansing and correcting any dirty and missing or erroneous data. You have also learned about item sets and association rules and their applications.

In this unit, you will explore methods for grouping variables and observations, discuss methods and applications for these, and summarize what your group has done thus far on your Vila Health project. As in the previous units, you will keep working on your individual project, as well, incorporating the concepts covered in each unit into your project as you learn them. This unit's assignment is related to Vila Health, and you and your group should focus on completing your Midpoint Review for your Vila Health project, but do not lose sight of your individual project and how grouping might be used there, as well.

Learning Activities

u05s1 - Studies

Readings

From the Capella library, read:

- In Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons:
 - Chapter 6: Descriptive Modeling.
 - Chapter 7: Interpreting Descriptive Models.

u05s2 - Study Group Tasks

The recommendations that your group agrees to this week need to be included as part of the Midpoint Review and Final Report and will contribute to your final group project grade. Therefore, it is critical that you do not skip any of these discussions. You will be able to make changes to your recommendations at any time during the course, but skipping a topic will negatively impact your grade.

Review the Analytics Internship: Applied Forecasting scenario. Consider the topics for discussion as you review the scenario. Search for details that are specific to these topics and that will aid you in making recommendations to Vila Health.

This week's topics for discussion:

- How might you use the grouping methods covered in this unit to identify similar counties?
- How might you use the grouping methods covered in this unit to identify similar variables (and potentially further reduce the number of variables needed in the model)?
- What are the assumptions and considerations in using these methods for these purposes? Are these met, and if not, what do you need to communicate to the stakeholders for this analytics project?

Keep in mind the role you are playing within your group. You may find that some information is more useful to your role than others.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

u05a1 - Analytics Internship - Midpoint Review

Overview

For your Midpoint Review of the Vila Health assignment, you will be summarizing what your group has done so far within the study group meetings for your Vila Health virtual internship to define the business problem, identify the sources of data, evaluate the applicability and utility of the data in relation to the business problem, identify response variables to be predicted using the model, audit and profile the data, and cleanse and correct the missing or erroneous data. In addition, you will propose recommended variable and observation grouping methods based on what you learned in this unit.

Instructions

For this assignment, write a paper to summarize the work that you have done on your group Vila Health project to this point. This summary should include the following:

- Define the problem and the source of the data and explain how they align or do not align and what you will do to rectify any misalignment.
- Identify and define the response variables that will be predicted in your model.
- Audit and profile the data and identify any potential data issues. Include any outputs, graphs, et cetera from your data profiling and audit efforts.
- Address the data issues, including missing values and erroneous data. Include a specification and code for all data manipulation you had to complete in order to address these data issues.
- Describe the variable and observation grouping methods that your group recommends using for the Vila Health scenario.
- Reflect upon your experience with the group work for this project.
- Write a 4–6 page paper that addresses each of the points above, and include the SAS code written and used in this assignment as an attachment or appendix (not included in the 4–6 pages).
- Support your methods and choices with references, where appropriate.

Your assignment will be scored on the following criteria:

- Define the problem and the source of the data and explain how they align or do not align and what needs to be done to rectify any misalignment.
- Define the response variables to be predicted in the model.
- Audit and profile the data and identify any potential data issues and include any outputs and graphs from the data profiling and audit efforts.
- Address data issues, including missing values and erroneous data, and include a specification and code for all data manipulation completed in order to address these data issues.
- Describe recommended variable and observation grouping methods used for a business setting.

- Present the status of the initial methods and findings of an advanced data analytics and modeling project to stakeholders in a succinct and relevant manner.
- Reflect upon group work experience for an advanced data analytics and modeling project.

Refer to the Analytics Internship – Midpoint Review Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 4–6 typed double-spaced pages, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

[APA Style and Format](#)

u05d1 - Grouping

Discuss the difference between observations and variables and what it means to group each. Give an example of grouping variables and describe at least one statistical method or algorithm used for grouping variables. Next, provide at least one example of an application where we may want to group or cluster the observations, explain what grouping the observations would provide us over modeling or analyzing them without grouping them, and describe at least one statistical method or algorithm for grouping variables.

Response Guidelines

Respond to at least two other learners and share with them the portions of their initial post that you agree or disagree with, and why.

Course Resources

[Graduate Discussion Participation Scoring Guide](#)

Unit 6 >> Variable Reduction and Selection Methods

Introduction

When we have to interpret and explain a model and its business interpretation in layman's terms, it is imperative to ensure that our model is as simple, or parsimonious, as possible. For more complicated or behind the scenes applications, like machine learning and artificial intelligence, it may be possible to utilize more complex multi-layered models or transformations that are not easily translated into plain English. In either case, however, it still makes sense to keep the minimum number of variables in the model as possible, while maximizing the ability of the model to accurately predict the response variable(s) without overfitting.

In the previous units, you have learned about how to identify groupings or similar observations and variables. Using only one or a few variables from a particular group of similar variables is one way of reducing the number of variables in your model. After this point, we, as analysts, have to decide how to determine which variables to include in our model and which variables to exclude, including the variable selection method (forward, backward, stepwise, all possible regressions, or some other, more complex method) and also the measure or method for comparing performance of the model. At other times,

our business knowledge or that of our colleagues and teammates tells us that certain variables should definitely be included in a particular model, and, at times, we will have to force the use of this variable on top of, or in addition to, the other variable selection or reduction methods we decide to use.

In this unit, you will work on your individual project and continue working with your group to evaluate each of these methods and select one that you (and your group) determine is appropriate in each case. Note that the metrics and methods that your group decides to use may not be appropriate for, or the same as, those that you select to use for your individual project, and that is perfectly acceptable.

Learning Activities

u06s1 - Studies

Readings

From the Capella library, read:

- In Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN:
 - Chapter 8: Predictive Modeling.
 - Chapter 9: Assessing Predictive Models.

u06s2 - Study Group Tasks

The recommendations that your group agrees to this week need to be included as part of the Final Report and will contribute to your final group project grade. Therefore, it is critical that you do not skip any of these discussions. You will be able to make changes to your recommendations at any time during the course, but skipping a topic will negatively impact your grade.

Review the Analytics Internship: Applied Forecasting scenario. Consider the topics for discussion as you review the scenario. Search for details that are specific to these topics and that will aid you in making recommendations to Vila Health.

This week's topics for discussion:

- What are the benefits and pitfalls of each method for reducing and selecting the variables for this analytics and modeling project?
- What method or combination of methods is your group selecting to use for this analytics and modeling project?
- What problems might arise from your selected method, and what might you need to consider as you proceed with determining which variables should be included in your model?

Keep in mind the role you are playing within your group. You may find that some information is more useful to your role than others.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

u06a1 - Modeling Variables Report

Overview

When modeling and analyzing data with many variables, it is important to create the simplest model possible that still predicts your response variable as accurately as possible. There are several different ways to go about doing so, and determining which method to use can be driven by many different factors, including industry standards, organizational best practices, or even simply by personal preference.

It is important to:

- Evaluate all of your options for reducing the number of variables in your models to the minimum necessary.
- Select the variables that are most effective in predicting the response variables.
- Select the combination of variables that, as a group, is the most effective in predicting the response variables.

In this assignment, you will select and justify one method for doing each of these, and apply them to your individual project organizational problem and associated data.

Instructions

For this assignment:

- Write a paper of 4–5 pages (not including variable table and references) summarizing your organizational problem, any observation or variable groupings you have identified as relevant for your analysis, your model evaluation metric or metrics that you have selected, and the variable selection method you are recommending using.
- Support your selections and recommendations with references where appropriate.
- Finalize and describe your observation and variable groupings.
- Describe your data model's performance or comparison method.
- Describe your chosen variable selection method.
- Summarize your selections in a variable summary table that includes the variables that you are working with for your project, identifies response and explanatory variables, and notes any variables that you have specifically chosen to exclude or include in your model and justification of those choices. Include any other information you feel is relevant about each variable. Provide justification, with references where appropriate.

Your assignment will be scored on the following criteria:

- Describe observed and finalized variable groupings.
- Describe a data model's performance or comparison method.
- Describe a chosen variable selection method.
- Summarize data modeling selections and provide justification with references where appropriate.

Refer to the Modeling Variables Report Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 4–5 typed double-spaced pages, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

[APA Style and Format](#)

Unit 7 >> Selecting an Optimal Model

Introduction

In the previous unit, we discussed variable selection and reduction methods and what metrics or methods can be used to determine which variables to use or keep in the model and which to drop. Sometimes the metrics or methods used for variable selection and reduction are the same as those used to select an optimal model, and sometimes they differ. Whether they are the same or not, you need some overall metric or method to determine, across all types of models that you are evaluating, which one is optimal for your modeling situation and business scenario or problem.

In this unit, you will complete your model selection process for your individual project, and submit your final individual model report. You will also, with your group, evaluate optimal models for the Vila Health scenario and start finalizing your model for that project as well.

Learning Activities

u07s1 - Studies

Readings

From the Capella library, read:

- In Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons:
 - Chapter 10: Model Ensembles.

u07s2 - Study Group Tasks

The recommendations that your group agrees to this week need to be included as part of the Final Report and will contribute to your final group project grade. Therefore, it is critical that you do not skip any of these discussions. You will be able to make changes to your recommendations at any time during the course, but skipping a topic will negatively impact your grade.

Review the Analytics Internship: Applied Forecasting scenario. Consider the topics for discussion as you review the scenario. Search for details that are specific to these topics and that will aid you in making recommendations to Vila Health.

This week's topics for discussion:

- What method or methods has your group chosen to select the optimal model?
- Is your model selection metric or method in line with existing literature in this industry?
- Are there any assumptions or considerations that your group needs to keep in mind when using your selected metric or method?

Keep in mind the role you are playing within your group. You may find that some information is more useful to your role than others.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

u07a1 - Final Individual Model Report

Overview

In previous units, you considered multiple models for your individual business problem and related analytics and modeling. You selected a single method or metric to compare your options, and identified an optimal model for your specific business problem. It is very important to be able to explain the model and how it works to colleagues and stakeholders at the organization you have built the model for.

In this assignment, you will practice these skills and submit a final report on your individual model in non-statistical terms wherever possible.

Instructions

For this assignment:

- Write a paper of 4–6-pages (not including references and attachments or addendums) that summarizes your business problem, the source of the data used, the optimal model, and how you selected that model as optimal for your individual business problem.
- Include the SAS code created for your individual project as an attachment or addendum to your assignment.
- Explain the model fit evaluation method or metric that you chose to use in to compare the models you built, and explain how you determined that your model was the optimal choice for your individual business problem. Support your choice of model fit evaluation method or metric with references.

- Summarize and explain your optimal model in business or industry language. Try to avoid highly statistical language wherever possible in your model summary. Think of your audience as non-statistician stakeholders that have to approve the use of this model for addressing your individual business problem.

Your assignment will be scored on the following criteria:

- Recommend a model selection method or metric for an organizational problem.
- Summarize a model in business or industry terms, including appropriate considerations for its use and related assumptions or potential issues.
- Present the results of an advanced data analytics and modeling project to stakeholders in a succinct and relevant manner.

Refer to the Final Individual Model Report Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 4–6 double-spaced pages, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

[APA Style and Format](#)

u07d1 - Group Dynamics, Responsibilities, and Collaboration

Reflect on your group experience thus far:

- What challenges did the group face these past two weeks?
- What decisions do you think may have influenced the solution your group included in your Midpoint Review?
- How did your group decide on the objectives and problem of the case study itself?
- How do you think you impacted the group's decisions?
- What challenges do you feel you are facing as you work together to solve the problem?
- What recommendations did you share with your group?
- What feedback have you given and received this week?

Response Guidelines

Respond to at least two other learners who are not part of your group. Share with them the information in their initial post that you found most helpful in helping you understand the concepts.

Course Resources

[Graduate Discussion Participation Scoring Guide](#)

Unit 8 >> Model Scoring

Introduction

Not always, but many times, it is necessary for us to translate the predicted values of our model to something that is easier for others to use. This is especially true when the output of our model is a probability or likelihood of an event occurring, since not everyone understands or can accurately interpret probabilities. Often times, when creating models to predict a likelihood or probability, we will translate these probabilities into a score, like in

credit scores or insurance scores. Other times, we simply translate the output into ranks, as in models used to determine which customers to market to, for example. In these cases, it makes it very easy and "fool-proof" to determine the top 10,000 customers to market to, as we have translated them into ranks where we can just take the ranks from 1 to 10,000, and know that these are our top 10,000 customers. Each model and application is different, and requires a different scoring or ranking or other interpretation.

In this unit, you will compare methods for scoring observations for your individual and group projects, and will create, explain, and interpret a score for your individual project.

Learning Activities

u08s1 - Studies

Readings

From the Capella library, read:

- In Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN:
 - Chapter 12: Model Deployment.

u08s2 - Analytics Internship: Scoring



[Analytics Internship: Scoring](#)

[Transcript](#)

In this Vila Health activity, you will learn that the scope for the Vila Health project has been changed in ways that will affect the work and recommendations your team has been preparing. Your group will need to discuss what the changes mean and how they will impact the analytic approach. As you gather information from your mentor, consider what questions you and your team will need to address in order to make your final recommendations. Click **Analytics Internship: Scoring** to view the Vila Health scenario.

Course Resources

[Analytics Internship: Scoring](#)

u08s3 - Study Group Tasks

The recommendations that your group agrees to this week need to be included as part of the Final Report and will contribute to your final group project grade. Therefore, it is critical that you do not skip any of these discussions. You will be able to make changes to your recommendations at any time during the course, but skipping a topic will negatively impact your grade.

Review the Analytics Internship: Scoring scenario. Consider the topics for discussion as you review the scenario. Search for details that are specific to these topics and that will aid you in making recommendations to Vila Health.

This week's topics for discussion:

- How do the changes in your internship assignment change the problem you are being asked to solve?
- Can you use all or most of what you have done so far, or is there some of your work that will need to be scratched or redone?
- How do the results from the creation of a model differ from those of creating a score?
- What does your group need to consider as you work toward creating a score or scorecard?

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

[Analytics Internship: Scoring | Transcript](#)

u08a1 - Individual Scoring Specification

Overview

In your readings this unit, you learned about methods for deploying your model and for creating scores to translate that model into something more usable by non-statisticians. For this assignment, you will be selecting one method of model deployment or score or rank creation and implementing it for your individual modeling project, as well as explaining how it should be used, in non-statistical terms, to your colleagues and stakeholders.

Instructions

For this assignment:

- Write a paper of 4–6-pages (not including references and attachments or addendums) that summarizes your business problem, the source of the data used, the optimal model and the scoring, ranking, or other model deployment method you selected for your business problem.
- Include the SAS code created for this assignment as an attachment or addendum to your assignment.
- Explain the scoring, ranking, or other model deployment method that you chose to use for your individual project and explain how you determined that this method was the optimal choice for your individual business problem. Support your choice of scoring, ranking, or other model deployment method with references.
- Summarize and explain how to use your scoring, ranking, or model deployment method in business or industry language. Try to avoid highly statistical language wherever possible in your specification, and think of your audience as non-statistician stakeholders that have to approve the use of this model and related scoring, ranking, or model deployment methods for addressing your individual business problem.

Your assignment will be scored on the following criteria:

- Recommend a scoring, ranking, or other model deployment method for a specified model and business problem.
- Create a specification for a score or ranking method or scorecard and score a set of observations using the model deployment method.
- Present the results of an advanced data analytics and modeling project to stakeholders in a succinct and relevant manner.

Refer to the Individual Scoring Specification Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 4–6 double-spaced pages, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

[APA Style and Format](#)

u08d1 - Translating a Model to a Score

Compare and contrast at least three scoring methods, ranking algorithms, or ways to deploy the model you created for your individual project, and recommend one of them. Remember to summarize your business problem and data sources to provide your classmates with context around your comparison and recommendation. Be sure to support your recommendation with references where appropriate.

Response Guidelines

Respond to at least two other learners and share with them the portions of their initial post that you found interesting, and why.

Course Resources

Graduate Discussion Participation Scoring Guide

Unit 9 >> Special Topics: Geospatial Data

Introduction

Geospatial data is becoming more and more readily available to use for various analytical and modeling purposes. In order to use this type of data, we first need to understand what it is, what we need to consider when using it, and how to work with this type of data. In this unit, you will read about and summarize what geospatial data is, considerations for this type of data, and ways we can manipulate, handle, and use this type of data in statistical software and for analytics and modeling. You will also work with your group to complete and submit your final group technical report for your Vila Health internship project.

Learning Activities

u09s1 - Studies

Readings

From the Capella library, read:

- Kiryluk, K., Li, Y., Sanna-Cherchi, S., Rohanizadegan, M., Suzuki, H., Eitner, F., . . . Gharavi, A. G. (2012). [Geographic differences in genetic susceptibility to IgA nephropathy: GWAS replication study and geospatial risk analysis](#). *PLoS Genetics*, 8(6), e1002765.
- Lin, L. (2010). [An ecological study of children commuting to school](#) (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. 3443189)

u09s2 - Study Group Tasks

The recommendations that your group agrees to this week need to be included as part of the Final Report and will contribute to your final group project grade. Therefore, it is critical that you do not skip any of these discussions. You will be able to make changes to your recommendations at any time during the course, but skipping a topic will negatively impact your grade.

Review the Analytics Internship: Scoring scenario. Consider the topics for discussion as you review the scenario. Search for details that are specific to these topics and that will aid you in making recommendations to Vila Health.

This week's topics for discussion:

- Does your Vila Health analytics and modeling project include the use of geospatial data?
- Are there any special considerations that need to be made or portions of your work that have to be revisited as a result of what you have learned in this unit about geospatial data?

- What do you need to make sure to include as you complete your final group report, if you are writing it with the intention of another analyst being able to recreate your analysis?
- What are the absolute essentials for including in an executive/stakeholder presentation on your analytics and modeling project? How do you take something this technical and detailed and summarize it in a meaningful way in layman's terms in a presentation? You will need to do so in the next unit's assignment.

For this unit's assignment, you will be completing and submitting your final group report for your Vila Health analytics and modeling project. Your report should contain everything that would be necessary for another analyst to pick this up and fully recreate your analysis and modeling process for the same or similar business problem. (Think about the possibility of someone having to rerun this analysis on updated data 5 or 10 years from now, for example.)

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

[Analytics Internship: Scoring | Transcript](#)

u09a1 - Analytics Internship - Final Report

Overview

For the final report of the Vila Health assignment, you will be summarizing what your group has done within the study group meetings for your Vila Health virtual internship to define the business problem, identify the sources of data, evaluate the applicability and utility of the data in relation to the business problem, identify response variables to be predicted using the model, audit and profile the data, and cleanse and correct the missing or erroneous data. In addition, you will recommend and apply observational and variable grouping methods, variable reduction and selection methods, optimal model selection methods, model interpretation and scoring, ranking, or deployment of your model. This will be the final group technical report for your Vila Health internship project.

Instructions

For this assignment, write a paper to summarize the work that you have done on your group Vila Health project throughout this course. This summary should include the following:

- The work you submitted in your Midpoint Review, incorporating changes and updates as a result of instructor and teammate feedback, including the following:
 - Define the problem and the source of the data and explain how they align or do not align and what you will do to rectify any misalignment.
 - Identify and define the response variables that will be predicted in your model.
 - Audit and profile the data and identify any potential data issues. Include any outputs, graphs, et cetera, from your data profiling and audit efforts.
 - Address the data issues, including missing values and erroneous data. Include a specification and code for all data manipulation you had to complete in order to address these data issues.
- In addition to the following summaries of work that has been completed since the Midpoint Review:
 - Describe the variable and observation grouping methods that your group used for the Vila Health scenario, as well as the results of their application.
 - Describe the variable selection and reduction methods that your group used for the Vila Health scenario, as well as the results of their application.
 - Describe the method your group used to select the optimal model for the Vila Health problem.
 - Explain your optimal model and provide the equation and other details for your model as appropriate.
 - Describe the scoring, ranking, or other deployment method and include a related specification that other technical or analytical colleagues could use to interpret and utilize your score to implement in systems and/or make business decisions.
 - Reflect upon your experience with the group work for this project and how it may have changed or stayed the same since the Midpoint Review.
- Support for your methods and choices with references, where appropriate.
- All code and screenshots of your work, included as attachments or addendums. Remember that the audience for a final technical model report is other statisticians, coworkers, and technical and analytical colleagues, so you may be as detailed and technical as necessary.

Your assignment will be scored on the following criteria:

- Apply analytic skills to current organizational problems using various advanced data analytics and modeling methods and techniques.
- Explain visual analytic approaches used for advanced data analytics and modeling.
- Present the results of an advanced data analytics project to stakeholders in a succinct and relevant manner.
- Reflect upon group work experience for an advanced data analytics and modeling project.

Refer to the Analytics Internship – Final Report Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 6–10 double-spaced pages, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

[Analytics Internship: Scoring | Transcript](#)

[APA Style and Format](#)

u09d1 - Considerations With Geospatial Data

Use this unit's readings and your research in the Capella library to describe what is meant by the term *geospatial data*, and what types of data are included in this category. Provide at least three considerations that are unique to this type of data, and describe at least three methods and tools to effectively use and handle this type of data. Which, if any, could have been used in the final project to further support the model? Please explain.

Response Guidelines

Respond to at least two other learners and share with them the portions of their initial post that you found interesting, and why.

Course Resources

[Graduate Discussion Participation Scoring Guide](#)

Unit 10 >> Communicating Results and Operationalizing Models

Introduction

In this unit, you will work with your group to complete and summarize the work you have done to address the Vila Health business problem. You will be given an opportunity to reflect on the last 10 weeks of this course, and work together to present your group's findings in a final presentation.

Learning Activities

u10s1 - Studies

There are no assigned readings for this unit.

u10s2 - Personal Effectiveness Assessment

For this activity, you will review your peers—and they you—on personal effectiveness skills, based on contributions to the group discussions. Please complete one survey for each member of your group. The assessment addresses the following characteristics:

- Interpersonal Skills.
- Integrity.
- Professionalism.
- Initiative.
- Dependability and Reliability.
- Willingness to Learn.

At certain points in your educational journey, you will meet with your department head for a 1:1 interview. The reviewer will gather your peer's assessments and your own self-reflections and use the data from these, in conjunction with the insight obtained during the 1:1 interview, to assess your skills. Your department head will provide you with feedback specifically intended to help you improve upon these skills.

During your capstone course the capstone panel of professionals will conduct a final review of your personal effectiveness skills. This final review will be assessed and your scores will be calculated as part of your final grade.

Course Resources

Personal Effectiveness Assessment

u10a1 - Analytics Internship - Final Presentation

Overview

For your final group presentation of your Vila Health internship for this term, you will be summarizing at a high level, and in business and industry terms, what your group has done within the study group meetings and final project for your Vila Health virtual internship.

Instructions

For this assignment:

- Summarize the business problem your mentor presented for the Vila Health project.
- Summarize the advanced analytics and modeling methods that you and your group used to address the Vila Health problem. Include references in the presentation notes, where appropriate.
- Summarize the results of your Vila Health advanced analytics and modeling project, including visual representations, identifying important variables, and instructions or specifications on using the model and associated scores. Include references in the presentation notes, where appropriate.
- Summarize your experience with the group work for this project and reflect upon the beginning to the end of the project.
- Write a 6–10 slide presentation with speaking notes to present your summaries.

Your group presentation should be written in a way that all terms used are common knowledge or are defined in layman's terms, that non-statisticians can understand all explanations, and that the instructions on using the model and scores are clear, concise, and complete. The audience for this presentation is all levels of management and stakeholders.

Your assignment will be scored on the following criteria:

- Summarize a business problem for a specific context.
- Summarize the advanced analytics and modeling methods used to address a specific business problem.

- Summarize the results of an advanced analytics and modeling project, including visual representations, identification of important variables, and instructions or specifications on using the model and associated scores.
- Present the results of an advanced data analytics project to stakeholders in a succinct and relevant manner.
- Reflect upon group work experience for an advanced data analytics and modeling project.

Refer to the Analytics Internship – Final Presentation Scoring Guide for more details.

Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of presentation:** 6–10 slides with speakers notes, not including references or addendums.
- **Font and font size:** Arial, 10 point.

Course Resources

[Analytics Internship: Advanced Analytics and Data Modeling | Transcript](#)

[APA Style and Format](#)

[Analytics Internship: Applied Forecasting | Transcript](#)

[Analytics Internship: Scoring | Transcript](#)