



Credit Hours: 3

Contact Hours: This is a 3-credit course, offered in accelerated format. This means that 16 weeks of material is covered in 8 weeks. The exact number of hours per week that you can expect to spend on each course will vary based upon the weekly coursework, as well as your study style and preferences. You should plan to spend 14-20 hours per week in each course reading material, interacting on the discussion boards, writing papers, completing projects, and doing research.

Faculty Information: Faculty contact information and office hours can be found on the faculty profile page.

COURSE DESCRIPTION AND OUTCOMES

COURSE DESCRIPTION:

In this course, students will investigate various statistical approaches used for data mining analyses. The preparation of data suitable for analysis from an enterprise data warehouse using SQL and the documentation of results is also covered. A simple data mining analysis project is performed to reinforce the concepts.

COURSE OVERVIEW:

In this course, you will investigate various approaches for discovering actionable intelligence within existing databases of information. You will learn about various techniques in data preparation and data manipulation of data suitable for analysis from enterprise data warehouses using SQL. Next, you will discover the core data mining topics of association, classification, and clustering. Finally, you will conduct a practical data-mining analysis project, using SAS, ETL, and SQL product, to reinforce the concepts taught in the course.

COURSE LEARNING OUTCOMES:

1. Analyze how online analytical processing (OLAP) and data mining are utilized to obtain business intelligence (BI).
2. Compare the types of statistical approaches used for data mining of enterprise data warehouse databases.
3. Formulate the extract, transform, and load (ETL) processes used to refresh a data warehouse based on a Star Schema.
4. Appraise the purpose of denormalized relational database data stored in materialized views and the role of ad hoc queries.
5. Construct a simple data warehouse with appropriate denormalized data using SQL for input to a statistical analysis software.
6. Create a data mining analysis using analytical software.
7. Assemble the findings of a data mining analysis in a professional business-oriented manner.

PARTICIPATION & ATTENDANCE

Prompt and consistent attendance in your online courses is essential for your success at CSU-Global Campus. Failure to verify your attendance within the first 7 days of this course may result in your withdrawal. If for some reason you would like to drop a course, please contact your advisor.

Online classes have deadlines, assignments, and participation requirements just like on-campus classes. Budget your time carefully and keep an open line of communication with your instructor. If you are having technical problems, problems with your assignments, or other problems that are impeding your progress, let your instructor know as soon as possible.

COURSE MATERIALS

Required:

Elliot, A. C., & Woodward, W. A. (2015) *SAS essentials: Mastering SAS for analytics* (2nd ed.). Hoboken, NJ: John Wiley and Sons. ISBN: 9781119042198

Larose, D., & Larose C. (2015). *Data mining and predictive analytics* (2nd ed.). Hoboken, NJ: John Wiley and Sons. ISBN: 9781118991121

Required Technology:

Please note the section at the bottom of this syllabus on “SAS Resources and Information on Certification.”

NOTE: All non-textbook required readings and materials necessary to complete assignments, discussions, and/or supplemental or required exercises are provided within the course itself. Please read through each course module carefully.

COURSE SCHEDULE

Due Dates

The Academic Week at CSU-Global begins on Monday and ends the following Sunday.

- **Discussion Boards:** The original post must be completed by Thursday at 11:59 p.m. MT and peer responses posted by Sunday 11:59 p.m. MT. Late posts may not be awarded points.
- **Opening Exercises:** Take the Opening Exercise before reading each week’s content to see which areas you will need to focus on. You may take these exercises as many times as you need. The Opening Exercises will not affect your final grade.
- **Mastery Exercises:** Students may access and retake Mastery Exercises through the last day of class until they achieve the scores they desire.
- **Critical Thinking:** Assignments are due Sunday at 11:59 p.m. MT.

WEEKLY READING AND ASSIGNMENT DETAILS

MODULE 1

Readings

- Chapters 1-3 in *Data Mining and Predictive Analytics*
- Karkouch, A., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 73, 57-81.

Opening Exercise (0 points)

Discussion (25 points)

Mastery Exercise (10 points)

MODULE 2

Readings

- Chapters 1 & 2 and Appendix E in *SAS Essentials*
- Data munging is a lot of work, so we made it easier. (2016). *Big Data Quarterly*, 2(4), 35.

Opening Exercise (0 points)

Discussion (25 points)

Critical Thinking (50 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

Option #1: Installation of PostgreSQL

There are two parts to this assignment:

1. You will install the PostgreSQL database and load the Northwind database. (See the link at the bottom of the page to access the instructions.) Capture a screenshot of the Northwind database after installation, similar to the example in the installation instructions.
2. Research SQL in data mining processes in the CSU-Global library. Write a brief paper explaining the role of SQL in data mining and statistical analysis.

Assignment Deliverables:

- Northwind database screenshot
- A paper on the role of SQL in data mining and statistical analysis

Your paper must meet the following requirements:

- Include a screenshot of the SQL installation.
- Be 2-3 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two peer-reviewed, scholarly references. The CSU-Global Library is a great place to find these resources.

You will have the opportunity to install the ETL tool as Option 2 for the Critical Thinking Assignment in Week 3. Click on the file linked in the Module 2 folder for detailed installation instructions for PostgreSQL.

Option #2: Installation of Pentaho ETL

There are two parts to this assignment:

1. You will install Pentaho ETL. (See the link at the bottom of the page to access the instructions.) Capture a screenshot of the Pentaho Data Integration launch window after installation, similar to the example in the installation instructions.
2. Research the role of ETL tools in providing clean and purposely transformed data as part of data mining processes. Write a brief paper explaining the role of ETL in data mining and statistical analysis.

Assignment Deliverables:

- ETL tool installation screenshot
- A paper on the role of ETL in data mining and statistical analysis

Your paper must meet the following requirements:

- Include a screenshot of the ETL tool installation.
- Be 2-3 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two peer-reviewed, scholarly references. The CSU-Global Library is a great place to find these resources.

You will have the opportunity to install the SQL tool as Option 1 for the Critical Thinking Assignment in Week 3. Click on the file linked in the Module 2 folder for detailed installation instructions for Pentaho ETL.

Mastery Exercise (10 points)**MODULE 3****Readings**

- Chapters 10 -12 in *Data Mining and Predictive Analytics*
- Otten, S., Spruit, M., & Helms, R. (2015). Towards decision analytics in product portfolio management. *Decision Analytics*, 2(1), 1-25.

Opening Exercise (0 points)**Discussion (25 points)****Mastery Exercise (10 points)****Critical Thinking (50 points)**

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

Option #1: Installation of PostgreSQL

If you installed PostgreSQL in Week 2, you must choose Option 2 this week.

There are two parts to this assignment:

1. You will install the PostgreSQL database and load the Northwind database. (See the link at the bottom of the page to access the instructions.) Capture a screenshot of the Northwind database after installation, similar to the example in the installation instructions.
2. Research SQL in data mining processes in the CSU-Global library. Write a brief paper explaining the role of SQL in data mining and statistical analysis.

Assignment Deliverables:

- Northwind database screenshot
- A paper on the role of SQL in data mining and statistical analysis

Your paper must meet the following requirements:

- Include the screenshot from the SQL installation.
- Be 2-3 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two peer-reviewed, scholarly references. The CSU-Global Library is a great place to find these resources.

Click the file linked in the Module 3 folder for detailed installation instructions for PostgreSQL. Refer to the Critical Thinking Assignment rubric in the Module 2 folder for more information on the expectations for this assignment.

Option #2: Installation of Pentaho ETL**There are two parts to this assignment:**

1. You will install Pentaho ETL. (See the link at the bottom of the page to access the instructions.) Capture a screenshot of the Pentaho Data Integration launch window after installation, similar to the example in the installation instructions.
2. Research the role of ETL tools in providing clean and purposely transformed data as part of data mining processes. Write a brief paper explaining the role of ETL in data mining and statistical analysis.

Assignment Deliverables:

- ETL tool installation screenshot
- A paper on the role of ETL in data mining and statistical analysis

Your paper must meet the following requirements:

- Be 2-3 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two peer-reviewed, scholarly references. The CSU-Global Library is a great place to find these resources.

Click the file linked in the Module 3 folder for detailed installation instructions for Pentaho ETL. Refer to the Critical Thinking Assignment rubric in the Module 2 folder for more information on the expectations for this assignment.

Portfolio Milestone (50 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

Option #1: Northwind Data Mining and Statistical Analysis Project – Planning

The objective of the Portfolio Project is mining data from a data warehouse, which contains data from the Northwind database you constructed during the installation of PostgreSQL.

Review the Portfolio Project milestone requirements in Modules 3 and 6, and the Portfolio Project requirements in Module 8 for a better understanding of the entire effort.

Summary of Tasks for the Portfolio Project:

Data Warehouse:

- Create a data warehouse database, including the fact and dimension tables (star schema).
- Create the schema for each table.
- Populate the tables using either ETL (Pentaho) or SQL (PostgreSQL).

Preprocessing for SAS:

- Extract data from the data warehouse, creating a file for input into SAS. The format of the file is your choice. Ensure SAS University Edition accepts your selected format.

Statistical Analysis Using SAS:

- Import data created in the preprocessing step.
- Conduct statistical analysis using the appropriate statistics from each category:
 - Summary statistics
 - Classification
 - Clustering
 - Association
- Prepare an analysis report.

Milestone Deliverables:

- A detailed plan including the tasks, activities, and software requirements
- A brief description of any challenges you might face in completing the Portfolio Project

Your paper must meet the following requirements:

- Be 2-3 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two peer-reviewed, scholarly references. The CSU-Global Library is a great place to find these resources.

Refer to the Portfolio Project milestone rubric in the Module 3 folder for more information on the expectations for this assignment.

Option #2: Clothing Store Data Mining and Statistical Analysis Project – Planning

The objective of this project is mining data from a data warehouse, which contains data from the Clothing Store csv file supplied to the class.

The Clothing Store file contents are covered in greater detail in Chapters 29-31 in the textbook, *Data Mining and Predictive Analytics*. This data set is large, with over 28,000 records and over 50 fields. You may wish to trim the data in the csv file before moving forward.

Review the Portfolio Project milestone requirements in Modules 3 and 6, and the Portfolio Project requirements in Module 8 for a better understanding of the entire effort.

Summary of Tasks for the Portfolio Project:

Data Warehouse:

- Create a data warehouse database, including the fact and dimension tables (star schema).
- Create the schema for each table.
- Load the Clothing Store csv into a database (which you will need to create), including the tables and the schema, or retain the data in csv format.
- Populate the tables using either ETL (Pentaho) or SQL (PostgreSQL).

Preprocessing for SAS:

- Extract data from the data warehouse, creating a file for input into SAS. The format of the file is your choice. Ensure SAS University Edition accepts your selected format.

Statistical Analysis Using SAS:

- Import data created in the preprocessing step.
- Conduct statistical analysis using the appropriate statistics from each category:
 - Summary statistics
 - Classification
 - Clustering
 - Association
- Prepare an analysis report.

Milestone Deliverables:

- A detailed plan including the tasks, activities, and software requirements
- A brief description of any challenges you might face in completing the Portfolio Project

Your paper must meet the following requirements:

- Be 2-3 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two peer-reviewed, scholarly references. The CSU-Global Library is a great place to find these resources.

Refer to the Portfolio Project milestone rubric in the Module 3 folder for more information on the expectations for this assignment.

MODULE 4

Readings

- Chapters 3, 4, & 8, *SAS Essentials*
- Knezek, G., Christensen, R., Tyler-Wood, T., & Gibson, D. (2015). Gender differences in conceptualizations of STEM career interest: Complementary perspectives from data mining, multivariate data analysis and multidimensional scaling. *Journal of STEM Education: Innovations & Research*, 16(4), 13-19.

Opening Exercise (0 points)

Discussion (25 points)

Mastery Exercise (10 points)

Critical Thinking (70 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

Option #1: Database and Data Warehouse Creation and Database Connections

Select this option if you have decided to complete Option 1 for the Portfolio Project. The purpose of this assignment is to introduce you to the technologies and processes needed to complete your Portfolio milestone and final project.

There are three parts to this assignment:

1. Create a new database and new data warehouse in PostgreSQL.
2. Establish database connections between the Jigsaw Operational Database and Jigsaw Data Warehouse.
3. Write a brief paper explaining key learnings and how they impacted your plan created in the Portfolio Project milestone for Module 3.

Assignment Deliverables:

- Output screenshots from each transformation
- A paper describing key learnings from each of the major steps (database creation, database population, and ETL transformations), and how these learnings impacted your plan created in the Portfolio milestone for Module 3

Your paper must meet the following requirements:

- Be 2-3 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two peer-reviewed, scholarly references. The CSU-Global Library is a great place to find these resources.

Click on the file linked below for detailed instructions to complete Parts 1 and 2. Refer to the Critical Thinking Assignment rubric in the Module 4 folder for more information on the expectations for this assignment.

Option #2: Database and Data Warehouse Creation and Database Connections

Select this option if you have decided to complete Option 2 for the Portfolio Project. The purpose of this assignment is to introduce you to the technologies and processes needed to complete your Portfolio milestone and final project.

There are three parts to this assignment:

1. Create a new database and new data warehouse in PostgreSQL.
2. Establish database connections between the Jigsaw Operational Database and Jigsaw Data Warehouse.

3. Write a brief paper explaining key learnings and how they impacted your plan created in the Portfolio Project milestone for Module 3.

Note: See link at the bottom of the page to access the instructions.

Assignment Deliverables:

- Output screenshots from each transformation
- A paper describing key learnings from each of the major steps (database creation, database population, and ETL transformations), and how these learnings impacted your plan created in the Portfolio milestone for Module 3

Your paper must meet the following requirements:

- Be 2-3 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two peer-reviewed, scholarly references. The CSU-Global Library is a great place to find these resources.

Click on the file linked below for detailed instructions to complete Parts 1 and 2. Refer to the Critical Thinking Assignment rubric in the Module 4 folder for more information on the expectations for this assignment.

MODULE 5

Readings

- Chapters 13-15 in *Data Mining and Predictive Analytics*
- Wang, S., Jiang, L., & Li, C. (2015). Adapting naive Bayes tree for text classification. *Knowledge and Information Systems*, 44(1), 77-89.

Opening Exercise (0 points)

Discussion (25 points)

Mastery Exercise (10 points)

Critical Thinking (70 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

Option #1: Statistical Analysis for a Life Insurance Company

Your organization, a life insurance company, wishes to analyze data from a heart health study to determine how to structure life insurance policies for individuals deemed at high risk for premature death. You are tasked with developing a better understanding of the variables in the data set containing research on heart conditions. Management wants you to explore this data set to determine if the data is suitable for use in the next phase of their upcoming analytics project.

Use SAS University Edition to conduct these statistical tasks:

- Data exploration
- Summary statistics
- Distribution analysis

- Table analysis

Note:

- Statistical tasks are located under **Tasks and Utilities > Tasks > Statistics**.
- The data set can be found in Libraries > **SASHELP > HEART**.

Submit an analysis of each variable in the data set. Include any tables, histograms, or scatterplot graphs necessary to support your analysis. Also, include a recommendation as to the suitability of this data set for meeting your organization's business goal.

The final analysis report must meet the following requirements:

- Be 4-6 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least three peer-reviewed, scholarly references, and one citation from the course textbooks. You may also include references from credible sources in print and from the Internet. The CSU-Global Library is a great place to find these resources.

Refer to the Critical Thinking Assignment rubric in the Module 5 folder for more information on the expectations for this assignment.

Option #2: Statistical Analysis for a Baseball Company

Your organization, a baseball agency, wishes to analyze data from a player performance study to determine if past performance is a predictor of future performance. You are tasked with developing a better understanding of the variables in the data set containing research on player performance. Management wants you to explore this data set to determine if the data is suitable for use in the next phase of their upcoming analytics project.

Use SAS University Edition to conduct these statistical tasks:

- Data exploration
- Summary statistics
- Distribution analysis
- Table analysis

Note:

- Statistical tasks are located under **Tasks and Utilities > Tasks > Statistics**.
- The data set can be found in Libraries > **SASHELP > BASEBALL**.

Submit an analysis of each variable in the data set. Include any tables, histograms, or scatterplot graphs necessary to support your analysis. Also, include a recommendation as to the suitability of this data set for meeting your organization's business goal.

The final analysis report must meet the following requirements:

- Be 4-6 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion

- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least three peer-reviewed, scholarly references, and one citation from the course textbooks. You may also include references from credible sources in print and from the Internet. The CSU-Global Library is a great place to find these resources.

Refer to the Critical Thinking Assignment rubric in the Module 5 folder for more information on the expectations for this assignment.

MODULE 6

Readings

- Chapters 19-22 in *Data Mining and Predictive Analytics*
- Zakharov, K. (2016). Application of k-means clustering in psychological studies. *Tutorials in Quantitative Methods for Psychology*, 12(2), 87-100.

Opening Exercise (0 points)

Discussion (25 points)

Mastery Exercise (10 points)

Critical Thinking (100 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

Option #1: Statistical Analysis for an Automobile Research Firm

Your organization, a consumer automobile research firm, wishes to analyze data from a study of fuel economy among the major automobile models to determine how the variables in the data set correlate with fuel economy. You are tasked with developing a better understanding of the variables in the CARS data set. Management wants you to explore this data set to determine if the data is suitable for use in the next phase of their upcoming analytics project.

You are required to conduct two analyses for this assignment.

1. Statistical Analysis

- Use SAS University Edition to conduct these statistical tasks:
 - Summary statistics:
 - Use MSRP, Invoice, MPG-City, and MPG-Highway as your analysis variables.
 - Use Make as your classification variable.
 - Distribution analysis:
 - Use MSRP, Invoice, MPG-City, and MPG-Highway as your analysis variables.

Note:

- Statistical tasks are located under **Tasks and Utilities > Tasks > Statistics**.
- The data set can be found in **Libraries > SASHELP > CARS**.

2. Cluster Analysis

- Conduct the following cluster analysis task:
 - Cluster variables:

- Determine which variables, if any, appropriately cluster the variables to account for variability.
- Limit your analysis to 10 clusters.

Submit an analysis of each of the variables used (MSRP, Invoice, MPG-City, and MPG-Highway). Include any tables, histograms, or scatterplot graphs necessary to support your analysis. Also, based on the cluster variables analysis, which variables, if any, can function as cluster variables? Provide tables, histograms, and other graphs to support your conclusion.

The final analysis report must meet the following requirements:

- Be 4-6 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least three peer-reviewed, scholarly references, and one citation from the course textbooks. You may also include references from credible sources in print and from the Internet. The CSU-Global Library is a great place to find these resources.

Refer to the Critical Thinking Assignment rubric in the Module 6 folder for more information on the expectations for this assignment.

Option #2: Analysis of Earthquake Research for an Emergency Assistance Organization

Your company, an emergency assistance organization, wishes to analyze data from a study of earthquakes around the globe to determine if the longitude and latitude are accurate variables to justify more emergency services in certain locations. You are tasked with developing a better understanding of the variables in the QUAKES data set. Management wants you to explore this data set to determine if the data is suitable for use in the next phase of their upcoming analytics project.

You are required to conduct two analyses for this assignment.

1. Statistical Analysis

- Use SAS University Edition to conduct these statistical tasks:
 - Summary statistics:
 - Use Longitude, Latitude, Depth, and Magnitude as your analysis variables.
 - Distribution analysis:
 - Use Longitude, Latitude, Depth, and Magnitude as your analysis variables.

Note:

- Statistical tasks are located under **Tasks and Utilities > Tasks > Statistics**.
- The data set can be found in **Libraries > SASHELP > QUAKES**.

2. Cluster Analysis

- Conduct the following cluster analysis task:
 - Cluster variables:
 - Determine if Longitude and Latitude are viable candidates for cluster variables.
 - Limit your analysis to 10 clusters.

The final analysis report must meet the following requirements:

- Be 4-6 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Supported with at least three peer-reviewed, scholarly references, and one citation from the course textbooks. You may also include references from credible sources in print and from the Internet. The CSU-Global Library is a great place to find these resources.

Refer to the Critical Thinking Assignment rubric in the Module 6 folder for more information on the expectations for this assignment.

Portfolio Milestone (100 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

Option #1: Northwind Data Mining and Statistical Analysis – Data Warehouse

The purpose of this milestone assignment is to complete the tasks described below in preparation for your final project delivery.

1. Data Warehouse:

- Create a data warehouse database, including the fact and dimension tables (star schema).
- Create the schema for each table.
- Populate the tables using either ETL (Pentaho) or SQL (PostgreSQL).

2. Preprocessing for SAS:

- Extract data from the data warehouse, creating a file for input into SAS. The format of the file is your choice. Ensure SAS University Edition accepts your selected format.

You should use the plan formulated in Milestone 1 of Module 3 for the detailed steps you intend to follow.

For this milestone assignment, you are expected to submit:

- Screenshots of the populated data warehouse
- Star schema design, either a drawing or screenshot
- Row counts for the fact and dimension tables
- Brief description of your key learnings from completing this assignment

Your assignment must meet the following requirements:

- Be 2-4 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two scholarly journal articles (at least one of which is peer-reviewed). The CSU-Global Library is a great place to find these resources.

Refer to the Portfolio Project Milestone rubric in the Module 6 folder for more information on the expectations for this assignment.

Option #2: Clothing Store Data Mining and Statistical Analysis – Data Warehouse

The purpose of this milestone assignment is to complete the tasks described below in preparation for your final project delivery.

1. Data Warehouse:

- Create a data warehouse database, including the fact and dimension tables (star schema).
- Create the schema for each table.
- Load the Clothing Store csv into a database (which you will need to create), including the tables and the schema, or retain the data in csv format.
- Populate the tables using either ETL (Pentaho) or SQL (PostgreSQL).

2. Preprocessing for SAS:

- Extract data from the data warehouse, creating a file for input into SAS. The format of the file is your choice. Ensure SAS University Edition accepts your selected format.

You should use the plan formulated in Milestone 1 of Module 3 for the detailed steps you intend to follow.

For this milestone assignment, you are expected to submit:

- Screenshots of the populated data warehouse
- Star schema design, either a drawing or screenshot
- Row counts for the fact and dimension tables
- Brief description of your key learnings from completing this assignment

Your assignment must meet the following requirements:

- Be 2-4 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least two fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least two scholarly journal articles (at least one of which is peer-reviewed). The CSU-Global Library is a great place to find these resources.

Refer to the Portfolio Project Milestone rubric in the Module 6 folder for more information on the expectations for this assignment.

MODULE 7

Readings

- Chapter 23, *Data Mining and Predictive Analytics*
- Chapters 18 & 19, *SAS Essentials*
- Ait-Mlouk, A., Agouti, T., & Gharnati, F. (2017). Mining and prioritization of association rules for big data: Multi-criteria decision analysis approach. *Journal of Big Data*, 4(1), 1-21.

Opening Exercise (0 points)

Discussion (25 points)

Mastery Exercise (10 points)

Career Development Criteria Assignment (30 points)

Practitioner Experience

For this assignment, you will research potential career and internship opportunities that may include data analytics or data mining as a requested or required skillset.

Instructions:

- Copy and include some of the ads, if possible.
- As you look for these positions, make notes of any educational requirements or specific certifications are required or recommended.
- Include in your research, articles about the job duties and responsibilities of a position in data analytics or data mining.
- What did you see that really interested you in learning more about a career in data analytics or data mining?
- Did you find out anything about a career in data analytics or data mining that concerned you?

Submission Requirements:

- You will report your research in a 3-page paper that meets the following requirements:
- Includes your initial research findings on your opportunities in the data analytics or data mining field.
- Reflects and reports any personal goals or future steps you might take to gain a competitive advantage in the workplace.
- Formatted according to the CSU-Global Guide to Writing & APA.

MODULE 8

Readings

- Chapters 29-31 in *Data Mining and Predictive Analytics*
- Delias, P., Doumpos, M., & Matsatsinis, N. (2015). Business process analytics: A dedicated methodology through a case study. *EURO Journal on Decision Processes*, 3(3), 357-374.

Opening Exercise (0 points)

Discussion (25 points)

Mastery Exercise (10 points)

Portfolio Project (200 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

Option #1: Northwind Data Mining and Statistical Analysis Project – Planning

The objective of this Portfolio Project is mining data from a data warehouse, which contains data from the Northwind database that was constructed during your installation of PostgreSQL.

Below are the summarized tasks for this Portfolio Project.

Data Warehouse:

- Create a data warehouse database, including the fact and dimension tables (star schema).
- Create the schema for each table.
- Populate the tables using either ETL (Pentaho) or SQL (PostgreSQL).

Preprocessing for SAS:

- Extract data from the data warehouse, creating a file for input into SAS. The format of the file is your choice. Ensure SAS University Edition accepts your selected format.

Statistical Analysis Using SAS:

- Import data created in the preprocessing step.
- Conduct statistical analysis using the appropriate statistics from each category:
 - Summary statistics
 - Classification
 - Clustering
 - Association
- Prepare an analysis report.

Using your plan prepared in Module 3, Milestone 1, and leveraging the data warehouse and preprocessing steps in Module 6, Milestone 2, complete the tasks under Statistical Analysis Using SAS.

Your analysis report must include:

- An analysis of each variable in the data set
- An analysis to determine which variables could serve as appropriate classifier variables
- An analysis to determine if any variables are candidates for clustering
- An analysis to determine if any variables have associations
- Any tables, histograms, or scatterplot graphs necessary to support your analyses
- A recommendation as to the suitability of this data set for meeting your organization's business goal

Your project must meet the following requirements:

- Be 6-8 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least four fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least three peer-reviewed, scholarly references, and one citation from the course textbooks. You may also include references from credible sources in print and from the Internet. The CSU-Global Library is a great place to find these resources.

Refer to the Portfolio Project rubric in the Module 8 folder for more information on the expectations for this assignment.

Option #2: Clothing Store Data Mining and Statistical Analysis Project – Planning

The objective of this Portfolio Project is mining data from a data warehouse, which contains data from the Clothing Store csv file supplied to the class. The Clothing Store file contents are covered in greater detail in Chapters 29-31 of the *Data Mining and Predictive Analytics* textbook. This data set is large, with over 28,000 records and over 50 fields. You may wish to trim the data in the csv file before moving forward.

Below are the summarized tasks for this Portfolio Project.

Data Warehouse:

- Create a data warehouse database, including the fact and dimension tables (star schema).
- Create the schema for each table.
- Load the Clothing Store csv into a database (which you will need to create), including the tables and the schema, or retain the data in csv format.
- Populate the tables using either ETL (Pentaho) or SQL (PostgreSQL).

Preprocessing for SAS:

- Extract data from the data warehouse, creating a file for input into SAS. The format of the file is your choice. Ensure SAS University Edition accepts your selected format.

Statistical Analysis Using SAS:

- Import data created in the preprocessing step.
- Conduct statistical analysis using the appropriate statistics from each category:
 - Summary statistics
 - Classification
 - Clustering
 - Association
- Prepare an analysis report.

Using your plan prepared in Module 3, Milestone 1, and leveraging the data warehouse and preprocessing steps in Module 6, Milestone 2, complete the tasks under Statistical Analysis Using SAS.

Your analysis report must include:

- An analysis of each variable in the data set
- An analysis to determine which variables could serve as appropriate classifier variables
- An analysis to determine if any variables are candidates for clustering
- An analysis to determine if any variables have associations
- Any tables, histograms, or scatterplot graphs necessary to support your analyses
- A recommendation as to the suitability of this data set for meeting your organization's business goal

Your project must meet the following requirements:

- Be 6-8 pages in length, not including the cover and references pages.
- Follow the CSU-Global Guide to Writing & APA. Your paper should include an introduction, a body with at least four fully developed paragraphs, and a conclusion.
- Be clearly and well written using excellent grammar and style techniques. Be concise and logical. You are being graded, in part, on the quality of your writing. If you need assistance with your writing style, you can find many writing resources in the CSU-Global Writing Center.
- Be supported with at least three peer-reviewed, scholarly references, and one citation from the course textbooks. You may also include references from credible sources in print and from the Internet. The CSU-Global Library is a great place to find these resources.

Refer to the Portfolio Project rubric in the Module 8 folder for more information on the expectations for this assignment.

COURSE POLICIES

Course Grading

20% Discussion Participation
0% Opening Exercises
0% Live Classroom
8% Mastery Exercises
37% Critical Thinking Assignments
35% Final Portfolio Project

Grading Scale	
A	95.0 – 100
A-	90.0 – 94.9
B+	86.7 – 89.9
B	83.3 – 86.6
B-	80.0 – 83.2
C+	75.0 – 79.9
C	70.0 – 74.9
D	60.0 – 69.9
F	59.9 or below

IN-CLASSROOM POLICIES

For information on late work and incomplete grade policies, please refer to our [In-Classroom Student Policies and Guidelines](#) or the Academic Catalog for comprehensive documentation of CSU-Global institutional policies.

Academic Integrity

Students must assume responsibility for maintaining honesty in all work submitted for credit and in any other work designated by the instructor of the course. Academic dishonesty includes cheating, fabrication, facilitating academic dishonesty, plagiarism, reusing /repurposing your own work (see *CSU-Global Guide to Writing and APA Requirements* for percentage of repurposed work that can be used in an assignment), unauthorized possession of academic materials, and unauthorized collaboration. The CSU-Global Library provides information on how students can avoid plagiarism by understanding what it is and how to use the Library and Internet resources.

Citing Sources with APA Style

All students are expected to follow the *CSU-Global Guide to Writing and APA Requirements* when citing in APA (based on the APA Style Manual, 6th edition) for all assignments. For details on CSU-Global APA style, please review the APA resources within the CSU-Global Library under the “APA Guide & Resources” link. A link to this document should also be provided within most assignment descriptions in your course.

Disability Services Statement

CSU-Global is committed to providing reasonable accommodations for all persons with disabilities. Any student with a documented disability requesting academic accommodations should contact the Disability Resource Coordinator at 720-279-0650 and/or email ada@CSUGlobal.edu for additional information to coordinate reasonable accommodations for students with documented disabilities.

Netiquette

Respect the diversity of opinions among the instructor and classmates and engage with them in a courteous, respectful, and professional manner. All posts and classroom communication must be conducted in accordance with the student code of conduct. Think before you push the Send button. Did you say just what you meant? How will the person on the other end read the words?

Maintain an environment free of harassment, stalking, threats, abuse, insults, or humiliation toward the instructor and classmates. This includes, but is not limited to, demeaning written or oral comments of an ethnic, religious, age, disability, sexist (or sexual orientation), or racist nature; and the unwanted sexual advances or intimidations by email, or on discussion boards and other postings within or connected to the online classroom. If you have concerns about something that has been said, please let your instructor know.

SAS RESOURCES AND INFORMATION ON CERTIFICATION

Information about SAS Certification

Taking SAS certification exams help you validate your skills and increase your value to an employer. You can choose SAS certifications across many subjects, including programming, data management, and analytics, to name a few. For more information on SAS certification go here: https://www.sas.com/en_us/certification.html. All students, teachers, professors or staff associated with an academic institution qualify for 50% discount on all SAS certification exams. Please contact certification@sas.com to receive the discount code that will reduce the exam fee by 50% during the registration process.

Resources for Learning SAS

SAS Certification Prep Guides: https://www.sas.com/store/books/categories/certification-guide/cBooks-cbooks_categories-cbooks_categories_12-p1.html

Visit SAS Communities Visit our online sites to share and connect with other SAS users and build your SAS skills. Don't miss key communities including: SAS Certification, SAS Training, SAS Academy for Data Science, SAS Programming, New SAS User, SAS Analytics U and SAS Viya for Learners. <https://communities.sas.com/t5/Learn-SAS/ct-p/learn>