D212 – Data Mining II

Welcome, Study Plan, and Course Pacing Guide

last update: 01.04.23

NOT SURE WHERE TO START IN THE COURSE?? You are in the right place!
Your D212 instructor team has put together the following guide full of resources and course tips to help you get the most out of this course and to help you pass the course in the most efficient way possible! Check out the Course Guide in Course Search.

Welcome to D212 – Data Mining II!

Data Mining II adds vital tools to data analytics arsenal that incorporates unsupervised models. This course explains when, how, and why to use these tools to best meet organizational needs. The prerequisite for this course is D211 – Advanced Data Acquisition.

D212 STUDY PLAN OUTLINE

(with Tips and Extra Resources!)

This study plan outline is a result of lots of student feedback on what has helped your fellow Night Owls be successful!

Many students find that the most efficient way to complete this course is by working through the Performance Assessment course projects directly and researching new topics as you encounter them in the project.

There may be some aspects of the projects you can complete using skills you might remember from previous Data Mining courses, like acquiring data and preprocessing data. But there will likely be some aspects of the projects that will be more unfamiliar to you – like segmenting data observations into clusters, scaling data to ensure that all variables are ranged to equally contribute to dimensionality results, preparing data suitable for market basket analysis, or sorting the top three rules of association tables.

As you encounter a task in the project you aren't as familiar with, that is a good time to consult the course materials in DataCamp, the supplemental resources linked within the "Resource Library" link located in the Course Information section at the bottom of the course of study, the last page in this document for instructor provided resources for help in understanding new concepts and seeing examples of those new algorithms in action!

D212 RECOMMENDED STUDY PLAN

1. **Read all the requirements of the Performance Assessments**, including the Task Descriptions as well as the Project Rubrics and the Weblinks section for additional information.

    a. You can access the Performance Assessment by clicking on the View Task link in the Assessment section of your course page.

    

    b. After reading through all the information about the Performance Assessments, consider watching the ***Video Resources in Course Tips and in Course Search*** for more detailed explanations about each section of each task and a ***walkthrough of the pain points*** showing you how to demonstrate different data preprocessing and performing analysis.

2. **Work through TASK 1: Clustering Techniques Parts A through I, TASK 2: Dimensionality Reduction Methods Part A through G, and Task 3: Association Rules and Lift Analysis Part A through H of the Performance Assessments.**[2] For each section:

    a. Include the answers to any questions asked in the PA in your written project document, as well as a screenshot of any code you write for that section.

    PRO TIP: SAVE YOUR CODE FREQUENTLY!! Be sure to save a copy of your code to your written document after completing each section. You can do this by retyping your code into your document, or by emailing yourself a copy of your code and then pasting the code into your document.

    

    b. Save and submit your cleaned dataset for evaluation.
    c. Submit your codes as part of your Performance Assessment.
    d. Always resubmit all parts of your Performance Assessment after each revision.

3. Before you submit your Performance Assessment for the first time, be sure to read through each requirement to help you maximize your chances of passing the Performance Assessment on your first attempt!

4. Submit your Performance Assessment! 😊
    a. If you do not pass on your first attempt, don't panic! 😊 View your evaluator's feedback by clicking on the View Task link in the Assessment section of your course page, and then clicking on the Evaluation tab on the next screen. Read your evaluator's comments and note what they ask you to change in each section that it not marked 'Competent' in green.

b. If you have any questions about how to go about making the requested changes, read through D212 Video Resources for extra tips and resources, and of course **reach out directly to your course instructor if anything isn't clear!**

D212 COURSE PACING GUIDE

Many students are able to complete this course in six weeks. To meet this **45-day Challenge**, we suggest the following pacing:

- **Week 1**: Read the Performance Assessment (PA) Task 1, watch video resources in Course Tips and engage the Learning Materials—focusing on a Clustering Technique (KMeans or Hierarchal) in DataCamp. Additional Learning Materials are located in the "Resource Library" link within the Course Information section at the bottom, of your course of study, below the Assessment section.

- **Week 2**: Start working on the PA for task 1 and plan to submit Task 1 at the end of Week 2; Don't forget to either email or schedule time with your Course Instructor for clarifications if needed.

Take the D212 Data Mining II 45-Day Challenge!

Read Performance Assessment and Watch Video Resources in Course Tips

- Watch the videos contained within the "Video Resources" topic in the Course Tips, then complete remaining sections of the Performance Assessment in Task 1, ***create your project video***, proofread, make final edits on your project documents, and submit your first attempts for evaluation.

- Reserve time for project revisions as required following evaluation; consult your course instructor as needed!

- **Week 3:** Read the Performance Assessment (PA) Task 2, watch video resources in Course Tips and engage the Learning Materials—focusing on Dimensionality Reduction Technique (Principal Component Analysis) in DataCamp. Additional Learning Materials are in Resource Library under Course Information on D212 home page.

- **Week 4**: Start working on the PA for task 2 and plan to submit Task 2 by the end of *Week 4*; Don't forget to either email or schedule time with your Course Instructor for clarifications if needed.

- Watch the videos contained within the Video Resources topic in Course Tips, then complete remaining sections of the Performance Assessment in Task 2. ***No project video is required for task 2.*** Proofread and make final edits on your project documents and submit your first attempt for evaluation.

- Reserve time for project revisions as required following evaluation; consult your course instructor as needed!

- **Week 5:** Read the Performance Assessment (PA) Task 3, watch video resources in Course Tips and engage the Learning Materials—focusing on Pattern Discovery Technique (Market Basket Analysis) in DataCamp. Additional Learning Materials are in Resource Library under Course Information on D212 course of study page.

- Watch videos contained within the Video Resources topic in Course Tips, then complete remaining sections of the Performance Assessment Task 3, ***create your project video,*** proofread, make final edits on your project documents, and submit your first attempt for evaluation.

- **Week 6**: Reserve time for project revisions as required following evaluation; consult your course instructor as needed!

Where to Get Help … for Each Section of the Assessment!

(Student favorites are marked with a red arrow: ←)

Scenario:

Task 1:

- Don't use PCA in task 1

Task 2:

- Use either Scenario 1 or 2

Task 3:

- Focus on either Scenario 1 or 2
- Use required task 3 dataset from Data Sets and Associated Data Dictionaries. ←

D212 Requirements and Tutorials:

PRIMARY LEARNING MATERIALS: Datacamp and Resource Library in "Course Information: section of the Course of Study page

- **Task 1 Overview & Tutorial:**
  - D212 Performance Assessment Task 1: Part 1 & 2 - Clustering Analysis ←
  - D212 Performance Assessment Task 1: Part 3 & 4 - Clustering Analysis ←
  - D212 Performance Assessment Task 1: Part 5 & 6 - Clustering Analysis ←
  - KMeans Presentation Background ←
  - Constructing and Fitting KMeans on Dataset ←
  - Evaluate KMeans Model ←
  - Analyze and Interpret KMeans ←

- **D212 Task 2: Overview & Tutorial:**
  - Data Mining II - D212 Performance assessment Overview ←
  - PCA with Scree Plot Elbow Method and Kaiser Rule Method ←
  - How to create a PCA Scree Plot Using Automated Process ←
- **D212 Task 3: Overview & Tutorial:**
  - Market Basket Analysis in Theory ←
  - Market Basket Analysis in Python ←
  - How to perform Market Basket Analysis in R ←
  - How To Install mlxtend in Anaconda Environment ←