

## D206 – Data Cleaning

### Welcome, Study Plan, and Course Pacing Guide

last update: 2.26.23



#### NOT SURE WHERE TO START IN THE COURSE?? You are in the right place!

Your D206 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! 😊

## Welcome to D206 – Data Cleaning!

Data Cleaning continues building proficiency in the data analytics life cycle with data preparation skills. This course addresses exploring, transforming, and imputing data as well as handling outliers. Learners write code to manipulate, structure, and clean data as well as to reduce features in data sets.

### D206 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 course resources before beginning the work on the Performance Assessment.



As you encounter a task in the project you aren't as familiar with, that is a good time to consult the course materials and the supplemental resources shared below, for help in understanding new concepts.

### D206 RECOMMENDED STUDY PLAN

The study plan below has helped many students succeed in this course.

1. **Get familiar with the course format and the performance assessment:** Watch [Webinar 1 "Getting Started with D206"](#), which can be accessed by selecting *Course Tips* on the right-hand side of the course page, then select *View All*.
2. **Study Data Cleaning Concepts:** This course contains two primary learning resources--WGU Courseware Materials and DataCamp Videos. The course also



contains supplemental resources (e.g., instructor-led videos). Review the [Course Resource Guide](#) which guides students on how to utilize all course resources, collectively.

- a. **WGU Courseware Material: Review Lessons 1-7**
  - b. **DataCamp Videos:** DataCamp resources can be found by selecting the “Go To Course Material” on the course page > Learning Resources.
  - c. **D206 Instructor-Led Webinars / Videos:** These videos and PowerPoints can be accessed by selecting *Course Tips* on the right-hand side of the course page, and then selecting *View All*.
    - [Webinar 2](#) – Getting Started with Missing Data and Outliers
    - [Webinar 3](#) – Getting Started with Re-expression of Categorical Variables
    - [Webinar 4](#) – Getting Started with PCA
  - d. **Select your programming language, either Python or R:** Review [R or Python](#) for some side-by-side comparisons. (You will continue to learn both R and Python)
3. **Select one of the available data sets:** The data sets can be found by selecting *View Task* on the course page and then selecting the *Data Sets and Associated Data Dictionaries* link under *Introduction*.
  4. **Complete the Performance Assessment:** After reading through all of the Performance Assessment requirements, consider reviewing the [“Performance Assessment Success Tips Guide”](#) for more detailed explanations about each requirement of the performance assessment task. This document can be found by selecting *Course Tips* on the right-hand side of the course page, then selecting *View All*.

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!

- 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 Assessments 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 make the requested changes, read through *D206 Video Resources* for extra tips and resources, and of course, **reach out directly to your course instructor if anything isn't clear!**

**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.

If you need additional help: call, email or set up a time to meet with your course instructor, or ***use the list of supplemental resources for each section on the following pages! <<<***

## D206 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:

The amount of time any course will take you to complete depends on many factors, including your background, previous experience with the course material, and the amount of time you can devote to studying. Many students can complete this course in 5-7 weeks. To help you stay motivated while making progress toward your goals, your course instructor group has put together a **45-day Challenge** for this course. To meet the challenge, we suggest following the recommended pacing:

**Take the D206  
Data Cleaning  
45-Day Challenge!**

**CHALLENGE  
ACCEPTED**

- **Week 1**: Familiarize yourself with the course resources and the task requirements. Choose your language, either Python or R.
- **Week 2**: Complete lessons 1-3 and the aligned Data Camp Videos.
- **Week 3**: Complete lesson 4 and the aligned Data Camp Videos.
- **Week 4**: Complete lessons 5 and 6 and the aligned Data Camp Videos.
- **Week 5**: Complete lesson 7 and the aligned Data Camp Videos.
- **Week 6**: Perform, write, and submit the Task. Work on revisions to the Task, if needed.

Your course instructors are here to help and encourage you with the course content. You can email or schedule an appointment with *any* course instructor if you have any questions about the course content. We want to help you succeed in this course!

You can create your own custom D206 pacing guide by using *Dr. Straw's D206 Pacing Guide CUSTOM* Excel file in [Dr. Straw's D206 resource folder](#).

## Where to Get Help ... for Each Section of the Assessment!

- *"Performance Assessment Success Tips Guide"* for more detailed explanations about this requirement of the performance assessment task.

**A**

### Part I: Research Question!

- *"Performance Assessment Success Tips Guide"*
- Dr. Straw's tips for success

**B**

### Part I: Variables!

- *"Performance Assessment Success Tips Guide"*
- Dr. Straw's tips for success

**C**

### Part II: Data Cleaning Plan! (Detection)

- *"Performance Assessment Success Tips Guide"*
- Webinar 2 – Getting Started with Missing Data and Outliers
- Webinar 3 – Getting Started with Re-expression of Categorical Variables
- Dr. Straw's tips for success
- WGU Course Ware: Lessons 5 & 6
- Data Camp Videos
- [Python or R?](#)

#### ***For Python***

- Lesson 5: Cleaning Data in Python (Chapters 1, 2 & 3)
- Lesson 7: Dealing with Missing Data in Python (All Chapters)

#### ***For R***

- Lesson 6: Cleaning Data in R (Chapters 1, 2, & 3)
- Lesson 8: Dealing with Missing Data (All Chapters)

## Where to Get Help ... for Each Section of the Assessment!

**D**

### Part III: Data Cleaning! (Treatment)

- ["Performance Assessment Success Tips Guide"](#)
- [Webinar 2](#) – Getting Started with Missing Data and Outliers
- [Webinar 3](#) – Getting Started with Re-expression of Categorical Variables
- Dr. Straw's tips for success
- WGU Course Ware: Lessons 5 & 6
- Data Camp Videos

#### For Python

- Lesson 5: Cleaning Data in Python (Chapters 1, 2 & 3)
- Lesson 7: Dealing with Missing Data in Python (All Chapters)

#### For R

- Lesson 6: Cleaning Data in R (Chapters 1, 2, & 3)
- Lesson 8: Dealing with Missing Data (All Chapters)

**E**

### Part IV: PCA!

- ["Performance Assessment Success Tips Guide"](#)
- [Webinar 4](#) – Getting Started with PCA
- Dr. Straw's tips for success
- WGU Course Ware: Lessons 7
- Data Camp Videos

#### For Python

- Lesson 9: Dimensionality Reduction in Python (Chapters 1 & 4)

#### For R

- Lesson 10: Unsupervised Learning in R (Chapter 3 Only)

**F**

### Creating Your Panopto Video!

- ["Performance Assessment Success Tips Guide"](#)
- Online Panopto resources:
  - [How to Create a Video Using Panopto](#)
  - [How to Submit Your Performance Assessment including Video Link](#)
- **Contact Assessment Services at 877-HELP-WGU Option 2 if you:**
  - Need Panopto access / no CREATE button
  - Can't adjust your video sharing settings / evaluator cannot access video
  - Can't find the D206 Student Assignment's folder to put your video into

**G**

### Using & Citing Sources!

- ["Performance Assessment Success Tips Guide"](#)
- Every source in section I must be cited within your document.
- [Student Writing Center](#)
  - [How to document sources](#) (article)
  - [How to Cite Webpages](#) (article)
  - [How to Cite Webpages](#) (video)

**H**

### Professional Communication!

- ["Performance Assessment Success Tips Guide"](#)
- [www.grammarly.com](http://www.grammarly.com)
- [WGU Student Writing Center](#)
- [I Need Help with Professional Communication](#), which includes links to Writing Center resources on writing, grammar, and more!