Course Syllabus

✎ **Edit**

DATA 115 (Fall 2022)

Introduction to Data Analytics

Course Details

Instructor: Priyanka Rao

Email Id: **priyanka.rao@wsu.edu (mailto:priyanka.rao@wsu.edu)**

Contact: Via email

Office Hours: Tuesday 1 - 2 pm, Thursday 4 - 5 pm Or via appointment.

Introductory Note

First of all, welcome to Data 115!

I hope I will get a chance to meet you in person sometime during your WSU journey. Secondly, sorry this is such a long document! Despite that, I hope you'll make time to read it through at least once during the first week of class as it contains a useful summary of the material we will be covering and will serve as a repository of important information and links. Finally, definitely feel free to reach out if you have any questions or concerns about the course or material, I'm always happy to chat about data, math, or academic life and would welcome opportunities to offer my perspective or simply serve as a sounding board. Welcome to WSU!

Course Overview

Welcome to DATA 115 - Introduction to Data Analytics!

The course describes DA Basic concepts, principles, and tools used in data analytics.

More specifically, this course provides an introduction to the field of data analytics. We will draw on recent and relevant materials from statistics, mathematics, computer science, artificial intelligence, and machine learning, as well as many application domains.

Motivated by natural questions that arise in simple data examples, we will cover many of the basic techniques for working with data including sourcing raw data, cleaning and processing, exploring and analyzing, and finally presenting conclusions. In order to provide a foundation for later courses in the major, we will also explore initial examples of many of the core topics that will be encountered. You will have plenty of opportunities to work with real data and the R programming language.

As befits a rapidly developing, interdisciplinary subject, we will draw on recent and relevant materials from statistics, mathematics, and computer science, as well as many application domains. Motivated by natural questions that arise in simple data examples, we will cover many of the basic techniques for working with data including sourcing raw data, cleaning and processing, exploring and analyzing, and finally presenting conclusions. In order to provide a foundation for later courses in the major, we will also explore initial examples of many of the core topics that will be encountered. You will have plenty of opportunities to work with real data and the R programming language.

In addition to familiarizing you with basic tools and methods, this course will provide a broad exposure to the diverse types of data analytics projects that are being conducted around the world. A key component of the course will be critically analyzing published data analytics works and discussing their strengths and shortcomings. Finally, as data-driven practices are becoming common in many career fields, we will focus on professional development topics such as presentation skills and examples of the ethical and legal issues that can arise in modern data analysis projects.

Remote Logistics

As this is the Global Campus section of the course, our course materials will be presented entirely online, mostly over Canvas, with some videos displayed over the Panopto system. We will also use a variety of web collaboration tools, including the Canvas forums, and the RStudio Cloud platform for computational examples. Access to these services will be discussed during the first week of classes. Generally, each week will present you with pre-recorded videos covering some of the course content along with lecture slides, notes, and assignments.

Course Materials

There is no required textbook for this course, as we will be making use of open source and online materials all semester. These three textbooks below are available as free .pdfs and will be referenced during lectures and assignments. In addition to these texts, shorter weekly readings will be posted to the course LMS. Data sets and programming scripts for the course will be uploaded to the LMS.

- Open resources:
  - **Introduction to Statistical Learning (https://hastie.su.domains/ISLR2/ISLRv2_website.pdf)** (James, Witten, Hastie, and Tibshirani)
  - **R for Data Science (https://r4ds.had.co.nz/)** (Wickham and Grolemund)
  - **ggplot2: elegant graphs for data analysis (https://github.com/hadley/ggplot2-book)** (Wickham)

Software

At the beginning of the course, we will focus on manipulating and extracting details from data using spreadsheet programs like Excel. Later segments of the course will provide training and examples using the R programming language, which you will install on your own computer or access through a cloud platform. R is an open-source language with a robust ecosystem of packages for data analytics that has become one of the most common languages for a broad range of data-related tasks.

No previous programming experience or knowledge of statistical software tools will be assumed. We will start with the basics of R in Week 3 and build up familiarity with standard data analytics libraries like dplyr, ggplot2, knitr, tibble, and tidyr.

Expectations for Student Effort

Students should expect to spend a minimum of 9 hours per week, engaged in the following types of activities: reading, listening to/viewing media, discussion, or conversation in the LMS or other academic technology, conducting research, completing assignments, and reviewing instructor feedback, studying for and completing assessments, etc. (Note this statement is required to be posted in all Global Campus courses).

Student Learning Outcomes (SLOs)

Students who successfully complete the course will be able to:

- Describe different types, uses, and structure of data sets
- Perform basic procedures to obtain, process (clean), and store data
- Understand and compute simple summary statistics and statistical models
- Construct simple scripts for processing, analyzing, and visualizing data
- Conduct exploratory data analysis
- Apply elementary (supervised and unsupervised) learning techniques
- Analyze published data analytics work across multiple application domains
- Professional Preparation
  - Work collaboratively on data analytics projects
  - Present data preparation processes and the results of analyses
  - Understand legal and ethical ramifications of data-driven projects
  - Curate and store data sets
  - Use common programming tools and computational platforms

Class communication

We will use the Canvas forums for course discussions. This is a great place to ask questions from your peers, as well as to get feedback on your ideas. Announcements and other official communications will be posted on Canvas platform as well as sent to your official WSU email accounts. You should check these messages

regularly to stay informed about upcoming due dates and updates to the syllabus. You can reach out to me via email. Please include "DATA 115" in the subject line for any messages concerning the course. I might read your email as soon as I receive it but it may take me a little while to respond. I will commit to responding within
24 hours but this does mean that queries sent immediately before a deadline may not receive substantive responses in time to be directly helpful, so please plan ahead.

Course Work

There will be five main types of graded assignments in this course.

You will have **6 Participation/Reading and Discussion Assignment (RDA) and 12 Homework Assignment (HWA).** Individual responses to the assignment will be due at midnight the following Tuesday. No late work will be accepted so it is better to submit something partial as opposed to nothing. Written responses for RDA must be submitted as .pdf files to Canvas. HWA is completed by online homework that has link.

Participation / Reading and Discussion Assignments (RDA) - 10%

These involve completing one or more short readings on a relevant data analytics topic. A set of discussion questions related to the reading will be provided and you will be required to make a discussion post responding to those questions, as well as writing summaries of the group discussion. These are more subjectively graded and will receive a score--the main goal here is participation.

No late work will be accepted but at the end of the semester your **<u>lowest one score will be dropped.</u>**

Homework Assignments (HWA) - 30%

Each Monday, a problem set will be assigned, covering the course material for the forthcoming week. These will usually be a mixture of direct questions about the lecture material and opportunities for you to apply the methods we discuss to real data. Individual responses to the assignment will be due at midnight the following Tuesday. No late work will be accepted but at the end of the semester your **<u>lowest two scores will be dropped</u>**. Written assignments must be submitted electronically as .pdf files through Canvas.

Midterm Exam - 40%

There will be a single midterm exam during the 8th week of class, covering the material from week 1 - 7. This exam will be open notes and submitted
electronically through Canvas.The exams are timed to be completed in 75 minutes or less.

Final Project - 20%

The final assessment in the course will be a project, where you will complete a data analytics task from beginning to end. More details will be discussed later in the semester but this will provide you with flexibility to tackle a topic of broader depth than those encountered in the weekly assignments. You will submitting a project report in week 15. Project topics will be finalized during Week 12 of the course.

Collaboration Policy

For weekly assignments, you are encouraged to work with other students in the class. However, the **work that you submit should be your own** and in particular, must be written in your own words and communicate your own understanding of the solution. If you do collaborate, please list the names of the other students you worked with on your submission. Obtaining solutions from external sources like Chegg or Coursehero for course problems will be considered a violation of the academic integrity policy with consequences described below.

Online Discussion Policy

The essence of education is exposure to diverse viewpoints. In your discussion posts, you'll meet students with vastly different opinions and backgrounds. You're encouraged to disagree with the substance of others' ideas and opinions but do so with an active sense of respect for one another, and without losing focus on the topic at hand. Personal attacks, inflammatory statements, flaming, trolling, and disruption of the discussion do not have a place in academic discourse. Postings must comply with University policy on the use of computing resources, including those regarding harassment and discrimination, as well as conform to the **WSU Community Standards . (https://communitystandards.wsu.edu/home/)**

Your instructors will promote high-quality academic discussions by removing any posts they view as disruptive of the educational process and alerting students whose posts have been removed that they have violated course expectations. Students who continue to misuse the discussion boards after a warning may be subject to removal of access rights, course failure, and referral to the Office of Community Standards. **Visit netiquette guidelines . (https://online.wsu.edu/currentstudent/netiquette-guidelines/)**

Grading

Assignment Breakdown

Assignment	Percent of Overall Grade
Participation	10%
Weekly Assignments	30%
Midterm	40%

Assignment	Percent of Overall Grade
Final Project	20%
Total	100%

Final Project — 20%
Midterm — 40%
Participation — 10%
Weekly Assignments — 30%

30% 10% 40% 20%

- Participation
- Weekly Assignments
- Midterm
- Final Project

Weekly — In-class — Project
10% — 20%
30%
40% — Midterm

Weekly Assignments 30%
Participation 10%
Final Project 20%
Midterm 40%
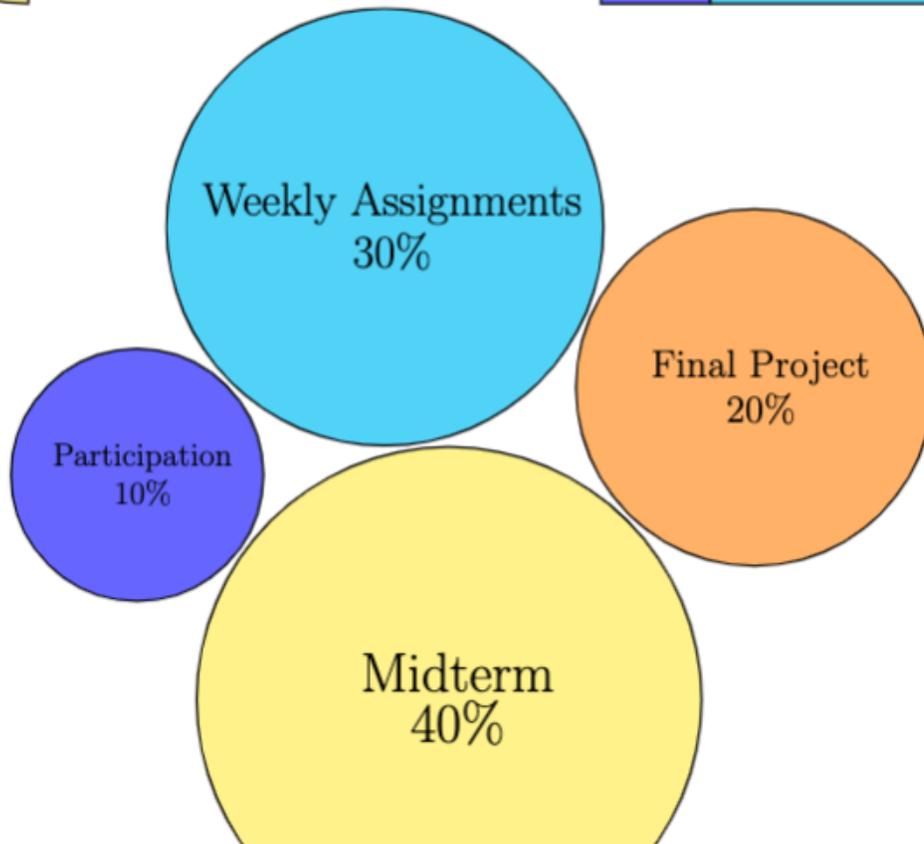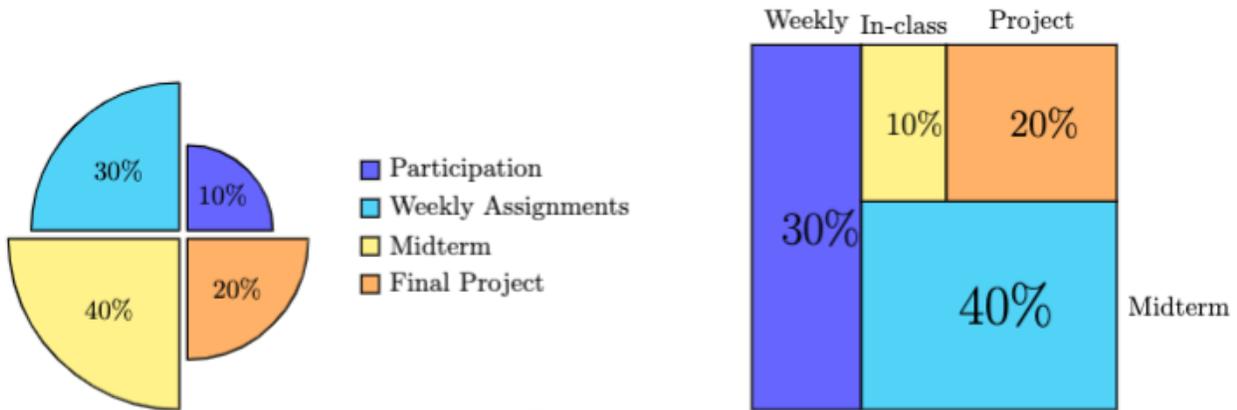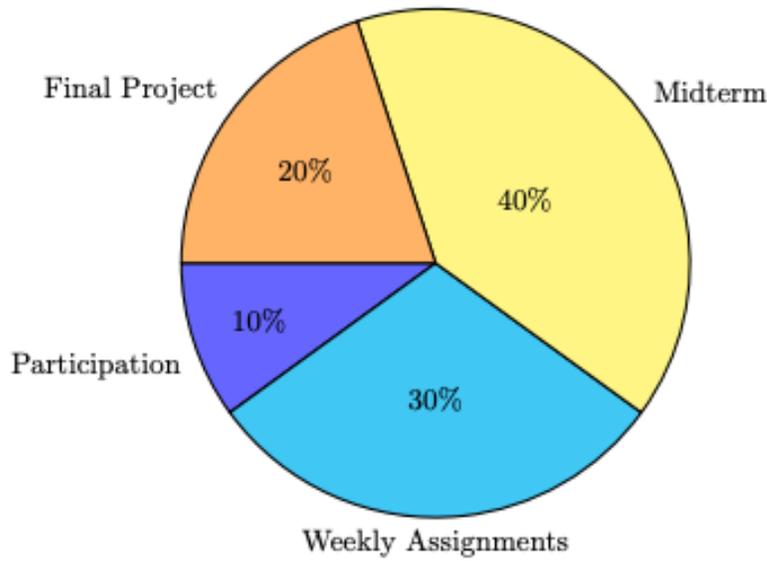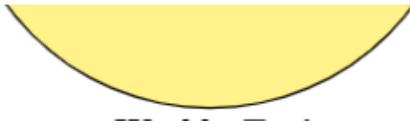
Grading Schema

Grade	Percent	Grade	Percent
A	95-100	C	73-76.99
A-	90-94.99	C-	70-72.99
B+	87-89.99	D+	67-69.99
B	83-86.99	D	60-66.99
B-	80-82.99	F	0-59.99
C+	77-79.99		

Instructor Interaction

We will use the Canvas forums for course discussions. This is a great place to ask questions from your peers, as well as to get feedback on your ideas. Announcements and other official communications will be posted on Canvas as well as sent to your official WSU email accounts. You should check these messages regularly to stay informed about upcoming due dates and updates to the syllabus.

I am accessible by email and text. Please include "DATA 115" in the subject line for any messages concerning the course.

Late Work Policy

Late Work Policy: Late submissions are not normally allowed. However, earlier submissions are allowed at any time before due date. Extensions *may* be allowed by contacting the instructor well in advance. After "answer keys" have been posted to the class, that assignment or exam cannot be made up, so communicate to the instructor and don't get too far behind. If asking an extension, do so on real emergencies—not as a habit.

It is better to turn in a partial homework than no homework. The phrase *"little is little but nothing is nothing"* is advisable here. Calculate the effect on your grade if you have all 9/10 on handful of assignments and for one you have a ZERO versus a SIX.

For an excused missed midterm exam, the accommodation is at the sole discretion of the instructor and may include a reweighting of the remaining components making up the student's grade or taking a

makeup exam. Again, contact the instructor well in advance.

Go to top of page

GC Include Placeholder(incompletePolicy)

GC Include Placeholder(creditHourEquivalent)

GC Include Placeholder(academicIntegrity)

GC Include Placeholder(Copyright)

GC Include Placeholder(academicRegulations)

GC Include Placeholder(DiscriminationAndHarassmentPolicy)

GC Include Placeholder(ReasonableAccommodations)

Go to top of page

GC Include Placeholder(AccommodationsForReligiousObservancesOrActivities)

GC Include Placeholder(SafetyAndEmergencyNotification)

GC Include Placeholder(onlineCollaboration)

GC Include Placeholder(WSUGlobalCampusStudentSupport)

GC Include Placeholder(librarySupport)

GC Include Placeholder(OnlineTutoring)

GC Include Placeholder(ImportantDatesAndDeadlines)

Go to top of page

GC Include Placeholder(courseSchedule)