

MIS458

Online Course (Asynchronous)
Data Wrangling
2023 Fall (Nov 28 2023 – Jan 13 2024)

- Instructor: Raymond (Buomsoo) Kim
- Phone: (515) 294-8198
- E-mail: rkim0927@iastate.edu
- Office: Gerdin 3220
- Office Hours: by appointment (Virtual)

Course Description

This course is designed to provide master's students in Management Information Systems (MIS) and related disciplines with practical knowledge and skills in acquiring, cleaning, transforming, and preparing data for analysis using Python programming language. Students will learn the fundamentals of data wrangling, explore various Python libraries and tools, and gain hands-on experience working with real-world datasets. In addition, they will learn basic concepts in data wrangling, along with recent trends in business analytics and data science. By the end of this course, students are expected to be proficient in data cleaning, manipulation, merging, and reshaping using Python.

Prerequisites

Students are expected to possess preliminary knowledge in programming and/or data management (such as database systems, data engineering, etc.). If a student is not certain, it is recommended to discuss with the instructor prior to the enrollment.

Learning Outcomes

1. Understand basic concepts in data wrangling and big data.
2. Understand the importance and role of big data in modern organizations.
3. Explore recent trends in business analytics and data science.
4. Gain proficiency in data cleaning and manipulation techniques using Python libraries.
5. Learn how to merge, reshape, and transform datasets using Python.
6. Apply data wrangling techniques to real-world datasets using Python.

Logistics

Course Format

This is a fully online course. The class will consist of multiple learning modules, which include:

- Videos (Uploaded Online)
- Python Courses, Tutorials, and Exercises
- Readings
- Online Discussions
- Group project

This is an entirely Web-based course. We have no regular face-to-face class meetings, and you are required to complete your work asynchronously - which simply means that you can work on it at different times than your colleagues. You can log into the class to do your work at whatever time is convenient for you as long as you are meeting the deadlines.

Nevertheless, it is important to understand that this is not a self-paced class or an independent study. You will have assigned deadlines, and work must be submitted on time. You may not save up your assignments to complete in the last weeks or days of the semester. One critical part of this class is regular interaction with other students and with me, your instructor. Each assignment sequence must be completed on schedule – you can't get behind and be successful.

Course Communication and Feedback

If you need to get in touch with me, the best method would be via email. Generally, I reply to emails within 48 hours and will provide feedback on assignments within 72 hours. You may also post questions pertaining to the course on the discussion board.

Announcements will be posted to this course whenever necessary. It is students' responsibility to regularly check the course website and ensure that your email account works properly in order to receive email.

When writing an email, include a header "<MIS548>" in the title and clearly articulate your concern in the body. Refer to the email template below if you are uncertain.

Title: <MIS548> Question regarding [insert question or topic]

Body:

Dear [insert instructor name],

My name is [insert first and last name]. I am a student in [class number].

I have a question about [insert question or topic].

Here is my work process towards the question [include your work or a picture of the work in the e-mail].

If you think it would be most helpful to meet virtually, I am available at these times over the next three business days.

[Meeting Option 1 with time zone]

[Meeting Option 2 with time zone]

[Meeting Option 3 with time zone]

Sincerely,

[First and Last Name]

Course Schedule

Please refer to a separate document on course schedule. It will be uploaded on course canvas page.

Course Requirements

Class Participation

Although this is an asynchronous, online course, students are strongly encouraged to actively participate. Students can participate by engaging in discussion boards – anyone can raise a question or provide an answer to questions posed by other students. Additionally, the instructor will have questions related to the assigned topics of the week. Both the quantity and quality of participation will matter.

Individual Assignments

There will be individual assignments during the semester that the students are required to complete before the deadline. The objective of the assignments is to check whether the student is following the course materials according to the given schedule.

- DataCamp assignments: Students are required to complete assigned Python assessments, which will be completed on DataCamp.

- End-of-the-semester reflection note: A reflection note (maximum 2 pages) summarizing what each student has experienced and learned during the course will have to be submitted by the end of the semester. A reflection note should include the following:
 - A student's thoughts on assigned articles and related topics. Students can share their thoughts on one of the topics or multiple topics throughout the course.
 - What a student learned by participating in a group project. Each student should describe in detail which role they played in the group and how they contributed to the project. In addition, they should elucidate what are the key lessons from the project, including those related to teamwork and data wrangling techniques.

Group Project

Students are required to complete one hands-on group project during the semester. Each group will consist of 3-5 students (depending on the total number of enrolled students). Role assignment and division of tasks within the group are encouraged. Regular progress updates and coordination should take place to ensure the project's smooth execution. At the end of the semester, all members in each group will be asked to report which role each member played and how much each member contributed to the team. These reports will be used to determine the peer evaluation scores.

Examples of roles that each student can be assigned include (but not limited to):

- Group leader
- Data engineer
- Business analyst
- Project manager
- Marketer

In the group project, students will have the opportunity to apply the data wrangling techniques and skills they have acquired throughout the course to a real-world dataset. The project aims to provide hands-on experience in tackling data quality issues, cleaning and transforming data, merging multiple datasets, handling missing data, and performing exploratory data analysis using Python. Each team is required to prepare a succinct report summarizing their data wrangling process and key findings.

- Project Tasks
 - **Dataset Selection:** Each group will select a real-world dataset relevant to business. The dataset should be sufficiently complex, requiring data

cleaning and manipulation.

- **Data Quality Assessment:** Conduct an initial assessment of the dataset to identify potential data quality issues, such as missing values, outliers, and inconsistencies.
- **Data Cleaning and Transformation:** Apply appropriate data cleaning techniques to address data quality issues. Transform and reshape the data as needed for analysis.
- **Data Integration:** Merge multiple datasets if applicable, ensuring compatibility and handling conflicts between different data sources.
- **Handling Missing Data:** Implement strategies to handle missing values effectively, such as imputation techniques or appropriate handling based on domain knowledge.
- **Exploratory Data Analysis:** Perform exploratory data analysis to uncover patterns, trends, and relationships within the dataset. Utilize Python libraries for data visualization to communicate the insights effectively.
- **Documentation and Presentation:** Prepare short video (maximum 12 minutes) summarizing your data wrangling process and key findings. In addition, a summary report (maximum 10 pages) documenting the details can be provided.

- **Assessment:**

The group project will be assessed based on the following criteria:

- Data quality assessment and identification of issues (10%)
- Effectiveness of data cleaning, transformation, and reshaping techniques (20%)
- Accuracy and efficiency of data merging and integration (15%)
- Handling of missing data and appropriate strategies used (10%)
- Depth and quality of exploratory data analysis (20%)
- Clarity and effectiveness of data visualizations (10%)
- Documentation and presentation quality (15%)

To prepare for a successful group project, each team is encouraged to actively participate in the following activities:

- **Weekly meeting with the instructor:** Although this is not mandatory, each team is strongly encouraged to participate in weekly meetings with the instructor to report their progress and obtain feedback. The weekly meetings will start in the second week and the instructor will provide several time slots so that as many team members can join the meeting as possible.

- **Group project milestone:** Each team is required to submit a milestone report by Week 3. The report should include each team's progress so far. It should not exceed 2 pages.

Evaluation and Grading Scale

Total score (in percentages) will be calculated as follows.

Category	Assignment	Type	Percent (%)
Class Participation	Online Discussions	Individual	15
	Participation in Group Project		10
	Total		25
Individual Assignments	DataCamp Assessment - Python Programming		5
	DataCamp Assessment - Data Manipulation with Python		10
	DataCamp Assessment - Importing & Cleaning Data with Python		10
	End-of-the-semester Reflection note		15
	Total		40
Group Project	Team Project Milestone Report	Group	5
	Team Project Final Report		30
	Total		35
Total			100

Grading scale

A final letter grade will be assigned based on percentages.

Points (percentages)	Letter Grade
90-100	A
80-89	B
70-79	C
60-69	D
Below 60	F

Late Work/Make-up Policy

All assignments are due by the deadline as posted on the course schedule.

Please plan accordingly and complete these assignments in advance of their deadlines to ensure any unanticipated circumstances do not result in a missed assignment. User error does not qualify you for any kind of makeup or retake opportunity.

You will be allowed to access the assignments an unlimited number of times until the due date/time indicated on the course calendar.

Late assignments will be subject to the following penalty: 20% will be deducted from the total score for every 24 hrs.

Academic Integrity & Grading Policy

You are expected to practice the highest possible standards of academic integrity. Any deviation from this expectation will result in a minimum academic penalty of your failing the assignment and will result in additional disciplinary measures. This includes improper citation of sources, using another student's work, and any other form of academic misrepresentation.

Use of Generative Artificial Intelligence (AI)

In all academic work, the ideas and contributions of others must be appropriately acknowledged. AI tools, including ChatGPT, are permitted in this course for certain assignments, and specific instructions will be included with each assignment. Identify any writing, text, or media AI generates when submitting work. Students should indicate how AI tools informed their process and the final product, including how AI-generated citations were validated. Failure to properly acknowledge the AI-generated contributions will be treated as academic misconduct.