

CS 185C: INTRODUCTION TO DATA ANALYSIS AND VISUALIZATION

Section 02, Spring 2021

San José State University
Department of Computer Science

COURSE AND CONTACT INFORMATION

Instructor:	Wendy Lee Ph.D.
Email:	wendy.lee@sjsu.edu
Class Days/Time:	Tue & Thu 12:00 pm – 1:15 pm (Pacific Standard Time)
Office Hours:	Schedule appointment @ www.sjsu.edu/people/wendy.lee/ Monday/Wednesday 10:00 am – 11:00 am
Classroom:	Online via Zoom
Prerequisites:	CS22B and graduate standing, or CS146 with a grade of C- or better

COURSE FORMAT

- This course will be conducted in a hybrid mode: Lectures and Labs will take place during live Zoom meetings. Zoom lectures will be recorded and posted in the Canvas Learning Management System at <https://sjsu.instructure.com>.
- Class time (live session) will be spent either in “lecture” mode or in “lab” mode, explained in “Class Protocol” in this document.
- Course materials such as syllabus, handouts, notes, hands-on exercises, projects, quizzes, exams, etc. can be found on Canvas Learning Management System. You are responsible for regularly checking with the Canvas messaging system to learn of any updates.
- Written and oral assessments will be used to measure student learning in this course.

COURSE DESCRIPTION

Topics in data wrangling, analysis and visualization. This course will cover tools and techniques in Python to efficiently work with and visualize large volumes of data in

meaningful ways to help solve complex problems in fields such as life sciences, business, and social sciences.

COURSE LEARNING OUTCOMES (CLOS)

Upon successful completion of this course, you will be able to

CLO1: Manipulate large datasets and handle missing or inconsistent values in datasets.

CLO2: Perform statistical analysis using numpy and scipy.

CLO3: Discover and visualize datasets using seaborn and matplotlib.

CLO4: Perform machine learning using Scikit-learn.

REQUIRED TEXTS/READINGS

Biological data exploration with Python, pandas and seaborn by Martin Jones. June, 2020. (<https://pythonforbiologists.com/biological-data-exploration-book>) ISBN-13: 979-8612757238

Hands-on Machine Learning with Scikit-Learn & TensorFlow by Aurélien Géon. March 2017. Publisher: O'Reilly Media, Inc. ISBN: 9781491962299

Additional course readings, examples, exercises, etc. will be assigned and provided by the instructor.

PYTHON PROGRAMMING ENVIRONMENT

We will be using Google Colab (<https://colab.research.google.com/>) with Chrome or any supported web browser, and program in Python within a Jupyter notebook. There is no additional software installation required.

COURSE REQUIREMENTS AND ASSIGNMENTS

1. **Quizzes (10%):** Quizzes will take place once a week at the beginning of class to assess students' knowledge of the course materials from the week prior. A unique password will be provided for each quiz during lecture.
2. **Hands-on Lab Report (40%):** The purpose of the hands-on lab is to develop your understanding of the material and your skills in problem-solving. You will work on the hands-on exercise with a group partner assigned by the instructor. Each student must write and submit independent lab reports. Hands-on lab

reports are only accepted in Canvas. You must submit lab reports on time to receive full credit.

3. **Midterms I & II (20%):** MT1 (10%): March 25, 2021, MT2 (10%): May 4, 2021. No make-up exams will be given if a student misses the midterm exam submission deadline (unless you have a legitimate excuse or other personal emergencies and can provide documented evidence).
4. **Final Project & Presentation (30%):** Final project (25%) and presentation (5%) will be used to assess student's understanding of the course materials at the end of the semester instead of a final exam. Each student will be given a unique problem to solve for the final project. The Final Project report and all the associated files must be submitted in Canvas are due on May 17, 2021 at 11:59 pm. Final Project Presentation needs to be pre-recorded and peer reviewed by classmates. Instructor will assign the presentations to each student for peer-evaluation. In order to get full credit for the presentation portion, everyone must evaluate the assigned presentations. Peer-evaluations are due on May 21, 2021 at 11:59 pm.

GRADING

Grading calculation will be based on the following:

- Quizzes - 10%
- Hands-on Lab Reports - 40%
- Midterm I & II - 20%
- Final Project - 30%

Late Submission – No late submission of lab report, exams, or final project will be accepted unless you have a documented emergency.

Exams: You may only submit your own work. Copying and any other form of cheating will not be tolerated and will result in a failing grade (F) for the course, as well as disciplinary consequences the university .

Grading Scale:

Point Range	Letter Grade	Point Range	Letter Grade
97.0 – 100	A+	72.0 – 76.99	C
93.0 – 96.99	A	70.0 – 71.99	C-
90.0 – 92.99	A-	67.0 – 69.99	D+
87.0 – 89.99	B+	62.0 – 66.99	D
82.0 – 86.99	B	60.0 – 61.99	D-
80.0 – 81.99	B-	<60.0	F
77.0 – 79.99	C+		

VIRTUAL CLASSROOM PROTOCOL

- **Live Session via Zoom:** Live Zoom meetings will be used as dual-purpose virtual classrooms. A meeting can be a regular lecture room, or it can be a computer laboratory for hands-on team exercises in break-out rooms.
- **Lecture Mode:** This is when Zoom is used as a virtual lecture room. You are expected to listen and follow the lecture. Be considerate to your classmates and follow the lecture. Keep your microphone muted except when speaking to the instructor. You may use the chat in Zoom to post questions during lecture.
- **Lab Mode:** Zoom break-out rooms will be used to group you into teams of three or more to work on hands-on lab exercises. Work collaboratively on the exercises and share your ideas and solutions with your classmates.
- **Attendance:** Live virtual class attendance is strongly encouraged.
- **Follow the rules of netiquette.** Be respectful. Dress appropriately if you are going to participate in the virtual classroom with the camera on.
- **Recording of Zoom Classes:** The instructor will record the live virtual classes using Zoom and the recordings will be shared in the Canvas course shell. If you do not wish to be identified in a class recording, please contact the instructor to arrange an “anonymous” option prior to class.
- **Zoom recordings and course materials:** You are allowed view the Zoom recordings for your own study purposes only. You may not record any course materials. You may not share any class recordings or course materials with someone who isn't enrolled in the without permission from the instructor. The lecture recordings and course materials are protected by copyright.
- **Accessibility:** If you need accommodations or assistive technology you should work with the Accessible Education Center (AEC) and the instructor.
- **Be Punctual:** Please arrive to the live sessions on time so that you benefit fully from the course experience and do not disturb classmates and the instructor while class is in session.
- **Stay on top of coursework:** You are responsible for knowing all material covered in lectures, assignments, quizzes, and course-related work.

TECHNOLOGY REQUIREMENTS

You are required to have an electronic device (laptop, desktop or tablet) with a camera and built-in microphone. SJSU has a free equipment loan program available for students: <http://www.sjsu.edu/equipmentcheckout>. You are responsible for ensuring that you have access to reliable Wi-Fi during tests. If you are unable to have reliable Wi-Fi, you must inform the instructor at least one week before the test date to determine an alternative. See [Learn Anywhere](#)

website: <https://www.sjsu.edu/learnanywhere/equipment/index.php> for current Wi-Fi options on campus.

VIRTUAL CLASSROOM ETIQUETTE

- Mute your microphone: To help keep background noise to a minimum, make sure you mute your microphone when you are not speaking.
- Be mindful of background noise and distractions: Find a quiet place to “attend” class, to the greatest extent possible.
- Avoid video setups where people may be walking behind you, talking, or making noise.
- Avoid activities that could create additional noise, such as shuffling papers or listening to music in the background.
- Position your camera properly: Be sure your webcam is in a stable position and focused at eye level.
- Limit your distractions/avoid multitasking: You can make it easier to focus on the meeting by turning off notifications, closing or minimizing running apps, and putting your smartphone away (unless you are using it to access Zoom).
- Use appropriate virtual backgrounds: If using a virtual background, it should be appropriate and professional.

UNIVERSITY POLICIES

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs’ Syllabus Information web page at <http://www.sjsu.edu/gup/syllabusinfo/>

COURSE SCHEDULE

The tentative course schedule is subject to change with fair notice. Changes will be announced on Canvas.

Week	Date	Topics
1	1/28	Syllabus. Introductions. Course Expectations. Google Colab, Jupyter
2	2/2	Introduction to Tables. <i>Hands-On #2</i>
2	2/4	Panadas, Series and dataframe objects. <i>Hands-On #3</i>
3	2/9	Panadas, Series and dataframe objects. <i>Hands-On #3</i>

Week	Date	Topics
3	2/11	Data exploration using pandas. <i>Hands-On #4</i>
4	2/16	Data exploration using pandas. <i>Hands-On #4</i>
4	2/18	Advanced features in pandas. <i>Hands-On #5</i>
5	2/23	Intro to seaborn & plotting special types of scatter plots. <i>Hands-On #6</i>
5	2/25	Plotting special types of scatter plots. <i>Hands-On #6</i>
6	3/2	Using Categorical axes with Seaborn. <i>Hands-On #7</i>
6	3/4	Using Categorical axes with Seaborn. <i>Hands-On #7</i>
7	3/9	Grouping and Categorizing data in pandas. <i>Hands-On #8</i>
7	3/11	Grouping and Categorizing data in pandas. <i>Hands-On #8</i>
8	3/16	Binning and ordered categories. <i>Hands-On #9</i>
8	3/18	Binning and ordered categories. <i>Hands-On #9</i>
9	3/23	Handling complicated data files. <i>Hands-On #10</i>
9	3/25	Midterm Exam #1
10	3/30	Spring Break - No Class
10	4/1	Spring Break - No Class
11	4/6	Matrix charts and heatmaps. <i>Hands-On #11</i>
11	4/8	Introduction to machine learning. <i>Hands-On #12</i>
12	4/13	Discover and visualize the data to gain insights. <i>Hands-On #12</i>
12	4/15	Prepare the data for Machine Learning algorithms. <i>Hands-On #13</i>
13	4/20	Introduction to Scikit-Learn. <i>Hands-On #14</i>
13	4/22	Feature scaling and Transformation pipelines. <i>Hands-On #15</i>
14	4/27	Select & Train a Model. <i>Hands-On #16</i>
14	4/29	Fine-tune the Model. <i>Hands-On #16</i>
15	5/4	Midterm Exam #2
15	5/6	Review Hands-ons 15/16
16	5/11	Review Hands-ons 15/16
16	5/13	Midterm 2 Answers
17	5/17	Final Project due.
17	5/21	Project presentations peer reviews due

Important dates:

02/08/2021: Last day to drop courses without an entry on student's permanent record.

02/15/2021: Last day to add courses for Spring 2021.

04/22/2021: Last day to Withdraw for Spring 2021.