

YSC2239 Introduction to Data Science

Semester 2, Academic Year 2021-2022

Time: TBA, Location: TBA

FACULTY NAME AND TITLE

Instructor: Michael Choi

Office Location: Elm RC2-M-05

Office Hours: TBA

Email: michael.choi@yale-nus.edu.sg

Zoom Office Hours: TBA

Zoom ID: 432 838 0728

Zoom Passcode: mchoi

Zoom Link: <https://yale-nus-edu-sg.zoom.us/j/4328380728?pwd=RmRhSGhLL1ZBVHBnbjdxTUUY3bm5Ndz09>

PEER TUTORS

Name: Nihal Zuhayar Parash Miaji

Email: nihalzuhayar@u.yale-nus.edu.sg

COURSE DESCRIPTION

This course revolves around the three major pillars of data science: mathematical and statistical thinking or modelling aided with computational tools. In the first part of the course, we will learn about basic programming in Python with focus in data science. Building upon these computational tools, in the second part of the course we shall cover topics about various statistical and data science methods. Throughout the course we will learn about concepts and methods such as p-values, hypothesis testing, central limit theorem, bootstrap, multiple linear regression, feature engineering, cross-validation and regularization, classification, logistic regression, decision trees and clustering, and familiarize ourselves with Python data science packages such as datascience, scikit-learn, numpy, scipy, pandas and matplotlib. Hands-on applications to real-world dataset are emphasized throughout the course, such as dataset in economics, geography, healthcare, manufacturing or e-commerce. The course culminates in a final group project in which students are expected to create appropriate data science models using the methods covered in class for data analysis.

COURSE GOALS

Upon completion of the course, the students are expected to:

- Use computing or statistical software, such as Python, to summarize data numerically and visually, and to perform data analysis.
- Model numerical response variables using a single or multiple explanatory variables.
- Select the most relevant explanatory variables in regression using statistical methods.
- Perform and evaluate data classification methods.
- Interpret results correctly, effectively, and in context without relying on statistical jargon.
- Critique data-based claims and evaluate data-based decisions.
- Complete an applied project that focuses on extracting insight from data using statistical inference or machine learning.

PRE-REQUISITES FOR THE COURSE

A level Mathematics or equivalent and YCC1122 Quantitative Reasoning or with the permission of the instructor. Prior experience in Python programming or knowledge in statistics or data science is helpful, but not required.

COURSE MATERIALS

We will mostly follow the two freely available e-book listed below and own notes of the instructor.

Title: Computational and Inferential Thinking: The Foundations of Data Science

Authors: Ani Adhikari and John DeNero

Link: <https://inferentialthinking.com/chapters/intro.html>

Title: An Introduction to Statistical Learning (ISLR)

Authors: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

Link: https://web.stanford.edu/~hastie/ISLR2/ISLRv2_website.pdf

COURSE ASSESSMENT BREAKDOWN

Attendance and participation 5%

Assignments 20%

Labs 15%

Project 8%

Peer evaluation 2%

Midterm 20%

Final 30%

DESCRIPTION OF ASSESSMENTS

Attendance and participation, Peer evaluation:

Students are expected to be present at all lectures, and actively participate in the discussion. The attendance and participation during these sessions, as well as activity on Piazza (the online discussion forum) will make up a portion of the grade in this class.

Up to two absences, regardless of reason, will be excused. All following absences will count towards your attendance/participation score.

The project of this class is team-based. Peer evaluation by team members will happen after the project presentations.

Assignments:

There will be weekly assignment on Canvas under Assignments in the first half of the course. The objective of the problem sets is to help students develop a more in-depth understanding of the material and help students prepare for exams, labs, and the project. Grading will be based on completeness as well as accuracy. Students must show all work in order to receive credit.

The lowest score will be dropped.

Students are welcomed, and encouraged, to work with each other on the problems, but each student must turn in his/her own work. If plagiarism is found, both parties will receive a 0 for the problem set grade as well as being reported to the Committee on Integrity and Discipline. Work submitted on Canvas will be checked for instances of plagiarism.

Submission instructions: Please submit the assignments in PDF or HTML format on Canvas. Please make sure that the file can be opened, otherwise it will receive 0 points. All assignments will be timestamped and late work will be penalized based on this time stamp. See late work policy below.

Labs:

There will be weekly Labs on Canvas under Labs. The objective of the labs is to give hands on experience with data analysis using Python. This is an individual-based assessment, and each individual should submit 1 HTML file on Canvas.

The lowest score will be dropped.

Similar to Assignments, students are welcomed, and encouraged, to work with each other on the problems in Labs, but each student must turn in his/her own work.

Project:

The course culminates in a final group project in which students are expected to create appropriate data science models using the methods covered in class for data analysis. Each group composes of 2-4 students, except in special circumstances approved by the instructor.

Further details on the projects will be provided as the due dates approach.

You must complete the project in order to pass this class.

Exams:

Midterm exam: TBA

Final exam: TBA

There will be one midterm and one final in this class. The midterm exam date is **tentative**, and the final exam date cannot be changed and no make-up exams will be given for both midterm and final. If you can't take the exams on these dates you should drop this class. You cannot pass the class if you do not take the final exam.

ABSENCE POLICY

Students are expected to attend all classes, and to notify the professor in advance if you will be absent. If you have required fieldwork/class trips for another course, you should discuss this potential absence with the instructor in advance.

LATE SUBMISSION POLICY

- Late work policy for assignments and labs:
Up to 24 hours late: Lose 30% of points
More than 24 hours late: Lose all points
- If you cannot complete an assignment on the due date due to a reported short-term illness with medical certificate or AD note, you have **until noon the following day** to complete it at no penalty. Then the regular late work policy will kick in.
- Late work policy for the project: 20% off for each day late.
- There will be no make-up for labs, assignments, project, or exams. If a midterm exam must be missed, absence must be officially excused **in advance**, in which case the missing exam score will be imputed using the final exam score, i.e. the final exam contributes 50% to the grade in this case. This policy only applies to the midterm. All other missed assessments will receive a grade of 0. The final exam must be taken at the stated time. You must take the final exam and turn in the project in order to pass this course.

- Regrade requests must be made within 3 days of when the assignment is returned, and must be submitted via email to michael.choi@yale-nus.edu.sg. There will be no grade changes after the final exam.
- Use of disallowed materials (textbook, class notes, web references, any form of communication with classmates or other persons, etc.) during exams and readiness assessments will not be tolerated. This will result in a 0 on the exam for all students involved, possible failure of the course, and will be reported to the Committee on Integrity and Discipline. If you have any questions about whether something is or is not allowed, ask me beforehand.

INTELLECTUAL PROPERTY AND PRIVACY

Our academic model encourages open and penetrating discussion of what can sometimes be challenging materials. Additionally, we seek to cultivate an intellectual space in which, as stated in the Faculty Statement on the Freedom of Expression, “there are no questions that cannot be asked, no answers that cannot be discussed and debated.” This kind of intellectual exploration requires trust and privacy. Therefore, students may not record and/or distribute course discussions, lectures, lecture slides or handouts, readings, videos, or any course related materials without prior permission of the instructor. This includes audio recording, video, transcription, and photography. Lectures and seminars that are delivered in-person, online, or as a pre-recorded videos should never be recorded or distributed beyond the course for which it was intended. Any notes which a student takes for their own learning and retention should not be shared beyond the Yale-NUS community. Students are encouraged to reflect upon and share their own learning experiences and ideas in whatever forum they wish. However, they should not share course content produced by their professors or their peers (e.g. a peer’s essay, comments made in class, posts to a Canvas discussion thread) without prior permission through any channels including social media.

Important Notes: Violation of this policy is addressed in the student Code of Conduct and could result in disciplinary and/or legal consequences. As per Clause F2(a) of Policies Relating To Yale-NUS College Intellectual Property, copyright to an Authored Work shall be owned by the University Member who authored it. Authored Work could include syllabi, tests, examination scripts, study guides, lecture notes and teaching materials, including lectures recorded on audio and/or visual recordings.

ACADEMIC INTEGRITY POLICY

Yale-NUS College expects its students to abide by the highest standards of academic integrity as a matter of personal honesty and communal responsibility. Acting with academic integrity requires that (a) students do their own work, (b) students not interfere with the work of others, (c) students accurately and honestly represent the content of their work, and (d) students properly attribute others’ work. Violations of the College’s academic integrity standards undermine both the community and the individual growth of students. Accordingly, they will be addressed with the utmost seriousness and sanctions ranging from grade penalties to expulsion. Examples of violations of academic integrity include plagiarism, copying or sharing homework answers, submitting work completed for one course as ‘new’ work for another course, or fabricating or falsifying research data. For more information please visit the Student Services website, Policies and Procedures section: <https://studentlife.yale-nus.edu.sg/policies/academic-integrity/>

The Yale-NUS Library provides resources on citations and plagiarism here: <http://library.yale-nus.edu.sg/avoiding-plagiarism/>

CLASS CLIMATE

Inclusive and Non-Discriminatory Language: Faculty and students will endeavour to learn and respect each other's names and preferred pronouns. While we aim to harness and learn from the diverse experiences and identities in our classroom, we will avoid expecting individual students to represent their entire country, culture, gender identity, etc.

Language & Conduct: This course encourages non-discriminatory language and conduct. Students should not use racist, sexist, ableist or other discriminatory language in class discussions or written work. Instead, students should be mindful and respectful of the diverse identities present in the class, including but not limited to gender, sex, sexuality, (dis)abilities, socioeconomic class, religion, race, nationalities, language. If you have suggestions to improve class climate and inclusivity, please come talk to me during office hours, send me an email, or submit anonymous feedback via Canvas. I understand it can be uncomfortable to talk to a professor about these issues, and I want to assure you that any concerns or constructive feedback you raise with me even if it is about my own speech or teaching style will not lead to negative consequences for your grade or our interactions in the course. If you feel uncomfortable speaking with me directly or want to consult with another faculty member about difficulties you are facing in the class, I encourage you to reach out to your Assistant Dean, who can advise you on the process for referring complaints relating to content or behaviour that causes offence to the College administration.

Access Needs: If you have any physical, psychiatric or learning conditions that may impact your performance in this course, please reach out to your Assistant Dean for advice and referral to further resources.

Class Discussions: Some students will be more comfortable or assertive speaking in class than others. Students who tend to speak more frequently, louder, and longer are encouraged to make room in the conversation for other voices and develop their listening skills. At the same time, students who are more reserved about speaking during class are encouraged to participate vocally or in other formats. Please come talk to me if participation or classroom discussion dynamics are challenging for you. We can work together to identify appropriate modes for participation.

If you are having inter-personal conflicts with a classmate in ways that undermines your learning or engagement, please come talk to me or talk to you AD so we can identify constructive ways forward.

HEALTH AND WELLNESS CONTACTS

Overall personal wellness is of the utmost importance. If you are experiencing undue stress and are concerned that your level of contribution to the class is being affected, please feel free to approach me. If you feel that you might benefit from private counselling, please contact your residential staff, Assistant Dean or the Yale-NUS Counselling Centre. For more information on the Counselling Centre, visit <https://studentlife.yale-nus.edu.sg/counselling/>

ACADEMIC RESOURCES

Canvas Page Usage Policy: All announcements and course material will be posted on Canvas. The students expected to check the course Canvas page regularly.

Research Consultations: Librarians at the Yale-NUS Library meet with students to assist them with their research and with developing information literacy skills and habits. Students should contact their Subject Librarian (<http://library.yale-nus.edu.sg/about/subject-librarians/>) or complete the Research Consultation form (<http://library.yale-nus.edu.sg/research-consultation/>) to set an appointment.

Writing Consultations and Peer Tutoring: The Yale-NUS Writers' Centre provides individual writing consultations for class assignments. For more information on making appointments, visit <https://writerscentre.yale-nus.edu.sg/> Many disciplines also have peer tutors available. To learn more and book an appointment, visit <https://teaching.yale-nus.edu.sg/peer-tutoring-programme/>

Schedule of Topics, Preparatory Assignments, and Assessments

See Canvas