

DATS 101

Introduction to Data Science

Subject to Change Statement

The syllabus and course schedule may be subject to change. Changes will be communicated via email and in NYU LMS (Brightspace). It is your responsibility to check email messages and course announcements to stay current in the course.

Course Information*

Course number: DATS 101

Course Title: Introduction to Data Science

Course Description:

This course provides an introductory exploration of data science, integrating probability and statistics, programming, and machine learning to prepare students for more advanced coursework. Designed for those with little to no background in machine learning, it offers a practical, hands-on approach that combines statistical reasoning, computational thinking, and real-world applications.

Unlike traditional machine learning courses that emphasize theoretical foundations, this course focuses on applied learning. Students will develop proficiency in Python and gain experience implementing data-driven models using NumPy, Pandas, and Scikit-learn. Topics will include fundamental probability and statistics concepts, core machine learning techniques, and essential programming skills for data analysis.

By the end of the course, students will have a strong foundational understanding of data science and be well-prepared to pursue more advanced topics in the field.

Co-requisite or prerequisite

- Introduction to Computer Programming or placed out ICP
- Probability and Statistics (or Honors Theory of Probability)

Course Fulfillments

- Data Science major requirement
- Algorithmic Thinking core requirement

Course Learning Outcomes

By the end of this course, students will:

1. Develop an understanding of fundamental probability and statistical concepts essential for data science.
2. Gain hands-on experience implementing core machine learning techniques, including regression, classification, and clustering.
3. Learn to write Python programs for statistical analysis, data visualization, and machine learning.
4. Develop the ability to assess and interpret model performance using appropriate evaluation metrics.
5. Apply data science techniques to real-world problems and datasets.

Course Requirements*

Class Participation

Class participation involves questions, discussion and comments during class and forums.

Assignments

There will be 7 homework assignments which are a mix of conceptual and programming-based assignments to reinforce theoretical understanding and practical skills. The written part of the homework assignments will be graded based on correctness and clarity in explaining the mathematical formula and conceptual results. The programming part of the homework assignment will be graded based on the correctness of coding giving the expected quantitative results, and coding clarity with clear algorithms and comments.

Mid-term and Final Exam

There will be one mid-term exam and one final exam which will be closed book. Exam questions will include mathematical derivation, concept verification, and problem-solving case studies.

Grading of Assignments

The grade for this course will be determined according to the following formula:

Assignments/Activities	% of Final Grade
Homework assignments	[30%]

Mid-term exam	[25%]
Final exam	[40%]
Class participation	[5%]

Letter Grades

Letter grades for the entire course will be assigned as follows:

Letter Grade	Points	Percent
A	4.00	95-100
A-	3.67	90-94
B+	3.33	87-89
B	3.00	83-86
B-	2.67	80-82
C+	2.33	77-79
C	2.00	73-76
C-	1.67	70-72
D+	1.33	67-69
D	1.00	60-66
F	.00	Below 60

View Grades

Grades will be available to the students after the due of each assignment on BrightSpace.

Course Schedule

The course is divided into three major components:

1. **Fundamentals of Probability and Statistics** – Covering probability theory, hypothesis testing, and statistical estimation methods.
2. **Foundations of Machine Learning** – Introducing key supervised and unsupervised learning models with practical applications.
3. **Programming for Data Science** – Integrated into recitation sessions, where students will implement key statistical and machine learning concepts in Python.

As a reference, here is [the syllabus](#) for Introduction to Data Science (DS-GA 1001) from NYU CDS.

The overall course structure will be as follows:

Topics and Assignments

Week/Date	Topic	Assignment
Week 1	Foundations of probability theory: events, random variables and probability distributions Python: random number generation	
Week 2	Random vector and important multivariate distributions	Homework 1
Week 3	Introduction to statistics, estimators, confidence interval, descriptive statistics Python: compute descriptive statistics of data	
Week 4	Hypothesis testing	Homework 2
Week 5	Hypothesis testing Python: implement hypothesis testing	
Week 6	Parametric inference	Homework 3

Week 7	Parametric inference	
Week 8	Byesian inference Python: generate samples from posterior distribution	Homework 4
Week 9	More reviews of linear algebra	
Week 10	Linear regression Python: implement simple and multiple regression	Homework 5
Week 11	Logistic regression Python: implement logistic regression on classification tasks	
Week 12	Regularization and model diagnostics	Homework 6
Week 13	Unsupervised Learning: Clustering Python: k-means implementation	
Week 14	Unsupervised Learning: PCA Python: implement PCA	Homework 7

Tests and Quizzes

1. Fundamentals of Probability and Statistics

- **Probability Concepts**
 - Foundations of probability theory: events, conditional probability, and Bayes' theorem
 - Random variables and probability distributions
 - Common distributions: Binomial, Gaussian (Normal), and Poisson
- **Statistical Inference**
 - Hypothesis testing: null and alternative hypotheses, p-values, and statistical significance
 - Significance tests: t-tests, chi-square tests, and ANOVA
 - Power analysis and confidence intervals
 - Randomized inference methods
 - Maximum Likelihood Estimation (MLE) and its applications.
- **Applied Statistics for Data Science**
 - Data summarization and visualization: descriptive statistics, histograms, and box plots
 - Correlation vs. causation: understanding confounding factors
 - Power analysis and experimental design: A/B testing and randomized controlled trials

Programming Integration:

- Writing Python functions for probability distributions
- Simulating random variables using NumPy
- Visualizing distributions and statistical results using Matplotlib and Seaborn

2. Foundations of Machine Learning

- **Regression Models**
 - Linear regression: Intuition, implementation, and evaluation
- **Classification Techniques**
 - Logistic regression for classification
 - Decision boundaries and model evaluation (precision, recall, F1-score)
 - Decision trees and their interpretability
- **Unsupervised Learning**
 - Clustering algorithms: K-Means and hierarchical clustering
 - Dimensionality reduction: Principal Component Analysis (PCA)
- **Model Selection and Evaluation**
 - Overfitting and underfitting
 - Cross-validation and hyperparameter tuning

Programming Integration:

- Implementing linear and logistic regression using Scikit-learn
- Performing classification and clustering tasks in Python

- Evaluating model performance with precision-recall curves and cross-validation

Course Materials

Recommended Textbooks and Resources

- Grus, J. (2019). *Data Science from Scratch: First Principles with Python*. O'Reilly Media.
- Wagh, S., Bhende, M., & Thakare, V. (2021). *Fundamentals of Data Science*. Wiley.
- Wasserman, L. *All of Statistics: A Concise Course in Statistical Inference*.
- DeGroot, M., & Schervish, M. *Probability and Statistics* (4th Edition).

Additional Notes

- This course serves as a replacement for **Mathematical Statistics and Econometrics**, incorporating additional programming components using data science libraries to provide a more applied and computational approach.

Course Policies

Attendance and Tardiness

You are expected to attend *all* scheduled classes. If you are unable to attend a class, notify me *before* that class.

Absence Exceptions

Observance of Religious Holidays: You may miss class for the observance of religious holidays. If you anticipate being absent because of religious observance, notify me in advance so we can create a plan for making up missed work. For more on this policy:

<https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/university-calendar-policy-on-religious-holidays.html>

Competitions, Conferences, Presentations: You are permitted to be absent from classes to participate in competitions, conferences, and presentations, either at home or out of town, as approved by the Associate Provost for Academic Affairs. Review the Undergraduate Bulletin for the conditions you must meet to obtain approval for this kind of absence.

Extended Illness: If you are ill and need to miss more than one week of classes, you must speak to the Health & Wellness Office. Once Health and Wellness verifies the reasons for your extended absence, I will consult with Academic Affairs and recommend the best course of action. I will not look at doctors' notes, both for your health privacy reasons and because I cannot verify the authenticity or content of the notes.

Tardiness

Punctual arrival is mandatory for this class. Be on time. Please do not leave in the middle of class unless it is an emergency.

Late Assignments

Assignments are due **at the date and time indicated on this syllabus**. Late homework assignments will not be accepted. Exceptions can be made only with instructor's prior approval.

Electronic Devices

Mobile Devices: You may not use mobile devices in class unless otherwise indicated.

Recording Class: To ensure the free and open discussion of ideas, you may not record classroom lectures, discussion and/or activities without my advance written permission; any such recording can be used solely for your own private use. If you have approved accommodations from the Office of Disability Resources permitting the recording of class meetings, you must present the accommodation letter to me in advance of any recording. Distribution or sale of class recordings is prohibited without the written permission of the instructor and other students who are recorded.

Instructional Technology

Email Communication: I will contact you regularly via the NYU Classes system. You should check for emails from me (reminders, logistics, updates, etc.) at the email address connected to NYU Classes. Please note that I will try to respond to your emails within 24 hours. Do not expect immediate responses to emails sent late at night or on the weekends.

Assignment Notification: All assignments will be posted to our class NYUClasses site. You are responsible for looking at NYUClasses after each class period to learn about the next homework assignment.

Instructional Tools: We will be using [fill in the blank] instructional technology resources in this class. You can learn about the tools used in this class through the [Student Digital Learning Toolkit](#). You are responsible for communicating with me in a timely fashion if you are facing any challenges using these tools.

Instructional Technology Assistance

If you need assistance with these tools, check the RITS website [here](#) or the RITS Toolkit [here](#), or you may email shanghai.rits@nyu.edu.

Academic Honesty/Plagiarism*

Carefully read NYU Shanghai's *Statement on Academic Integrity* (in the Undergraduate Bulletin). Breaches of academic integrity could result in failure of an assignment, failure of the course, or other sanctions, as determined by the Academic Affairs office.

Disability Disclosure Statement*

NYU Shanghai is committed to providing equal educational opportunity and participation for students with disabilities. It is NYU Shanghai's policy that no student with a qualified disability be excluded from participating in any NYU Shanghai program or activity, denied the benefits of any NYU Shanghai program or activity, or otherwise subjected to discrimination with regard to any NYU Shanghai program or activity. Any student who needs a reasonable accommodation based on a qualified disability should register with the [Moses Center for Student Accessibility](#) for assistance. Students can [register online](#) through the Moses Center and can contact the Academic Accommodations Team at shanghai.academicaccommodations@nyu.edu with questions or for assistance.

Title IX Statement

Title IX of the Education Amendments of 1972 (Title IX) prohibits discrimination on the basis of sex in educational programs. It protects victims of sexual or gender-based bullying and harassment and survivors of gender-based violence. Protection from the discrimination on the basis of sex includes protection from being retaliated against for filing a complaint of discrimination or harassment. NYU Shanghai is committed to complying with Title IX and enforcing University policies prohibiting discrimination on the basis of sex. Mary Signor, Executive Director of the Office of Equal Opportunity, serves as the University's Title IX Coordinator. The Title IX Coordinator is a resource for any questions or concerns about sex discrimination, sexual harassment, sexual violence, or sexual misconduct and is available to discuss your rights and judicial options. University policies define prohibited conduct, provide informal and formal procedures for filing a complaint and a prompt and equitable resolution of complaints.

Links to the Title IX Policy and related documents:

- [Sexual Misconduct, Relationship Violence, and Stalking Policy](#)
- [Procedures for Complaints Against Students](#)
- [Procedures for Complaints Against Employees](#)
- [Resource Guide for Students](#)
- [Resource Guide for Employees](#)

Academic Resources

ARC Services

The Academic Resource Center (ARC) offers both individual, one-on-one tutoring as well as group sessions in a variety of ways, in a variety of courses. You can log on to [WCOnline](#) to book an appointment with a Global Writing & Speaking Fellow or a Learning Assistant (LA). The Global Writing & Speaking Fellows conduct individual consultations on writing, speaking,

reading, and academic skills coaching. LAs provide both individual and small-group tutoring support in over 30 STEM, Business, Economics, IMA/IMB, and Chinese Language classes. Visit shanghai.nyu.edu/arc for more information about ARC services.

The Academic Resource Center (ARC) offers distance support for students who are enrolled in NYU Shanghai courses during the mixed teaching mode. Students taking classes at another campus/site should also take advantage of academic support resources available at those locations. You can log on to [WOnline](#) to book an appointment with a Global Writing & Speaking Fellow or a Learning Assistant (LA). The Global Writing & Speaking Fellows conduct individual, synchronous online consultations on writing, speaking, reading, and academic skills coaching. LAs provide both individual and small-group online tutoring support in over 30 STEM, Business, Economics, IMA/IMB, and Chinese Language classes. Visit shanghai.nyu.edu/arc for more information about ARC services and refer to [this site](#) to learn more about how to access online tutoring services.

Library Support

Library Services

The Library is available to support your research needs. They have access to over 20,000 print resources, 2,000 DVDs, and 1,000 databases (including over a million e-books, as well as streaming audio and video and image databases).

Librarians with expertise in your research topic are available to meet either *in person* or *online by appointment* or *by email* to help you navigate the research process, from developing a research question and formulating a research strategy, to selecting databases, requesting materials, and citing your sources. Visit shanghai.nyu.edu/library for more information on:

- 24/7 access to e-books, e-journals, streaming media, and databases
- Booking one-on-one consultations for research help
- [Asking the Library](#) questions via chat or e-mail

Electronic Reserves

Students can access course readings using their NYU credentials for courses they currently enrolled in at <https://ares.library.nyu.edu/>

Interlibrary Loan Service

For materials not available to you immediately, you can request scanned copies of a book chapter or journal article through our [Interlibrary Loan \(ILL\) service](#). If you don't know which chapter you need, you can request a Table of Content through ILL.