# Course Prefix Number – Syllabus CS 332

**INTRO TO APPLIED DATA SCIENCE WITH PROGRAMMING**

**Credits:** **4**

**Instructor Name: Ami Gates**

**Instructor Email:** **Ami.Gates@Oregonstate.edu**

**Teaching Assistant Name and Contact Info: TBD**

## Course Description [For Catalogue]

Explores the data science life cycle using appropriate Python libraries. Topics include current data science areas, data gathering, cleaning, formatting, and types, an introduction to supervised (decision trees) and unsupervised (clustering) learning, exploratory data analysis, data communication, decision science methodologies, storytelling, and narratives in data science.

## Course Description [For the Syllabus in the class]

The course will offer learners an overview of data science, data analytics, the data science life cycle, and using Python and common Python libraries/packages including Sklearn to perform data science and analytics tasks. The course will begin with a discussion and review of topics in current Data Science areas such as AI, Machine Learning, Deep Learning, Supervised and Unsupervised Learning, and related. Students will explore the stages of the Data Science Lifecycle beginning with methods for gathering data, such as the use of APIs via Python, methods for cleaning and preparing data, understanding, and identifying data types, data formats (such as record, transaction, sequential, image, and text, as well as labeled vs. unlabeled data), and concepts related to preparing data for various analyses. Students will engage in exploratory data analysis, as well as data transformation (such as normalization, discretization, etc.). Students will be introduced to unsupervised learning through k-means clustering as well as supervised learning though decision tree modeling. Students will use Python and appropriate Python Libraries to perform data cleaning and preparation, k-means clustering, decision trees modeling, and results illustration. Throughout this class, students will traverse and explore all stages of the Data Science Lifecycle. Students will also be introduced to, and will utilize, introductory data visualization methods and will engage in data communication, decision science methodologies, storytelling, and narratives in data science.

## Prerequisites or Corequisites

**Required:** CS 261

**Recommended:** MTH 252 or MTH 264. Python is preferred and recommended. Students should be prepared to code in Python/Anaconda during the class.

## Instructor Communication

Please post all course-related questions in the Q&A Discussion Forum so that the entire class may benefit from our conversation. Please contact me privately via OSU Email for matters of a personal nature. You can expect a reply to your questions within 24 hours on business days (Monday-Friday). Grading and providing feedback on your assignments and activities may take up to five days. If I need more time, I will let you know when to expect a response.

## Expectations for Time and Participation

This is an online, asynchronous, and partially flexible eCampus course. All course materials, requirements, resources, and submissions will take place in Canvas. Any synchronous materials, meetings, or video content will be recorded and posted in Canvas to enable student-selected review.

This course will require an average of 12 hours per week depending on previous experience with and knowledge of Python Programming. This course combines approximately 90 hours of instruction, online activities, and assignments for 4 credits.

Please note that this course, while online and partially flexible, is not self-paced. Our schedule of Course Content and the due dates that appear in Canvas provide guidelines for how you’ll interact and with what frequency. I recommend that you create your own workload schedule and set reminders for assignment due dates.

## Learning Resources

* This course provides all required materials at no cost to you. All materials are available within Canvas or free online.
* **Requirement:** This course requires the use of Python (Anaconda). Students may download and install Python/Anaconda for free and for PC, MAC, or Linux by going to this site and following the instructions: <https://www.anaconda.com/download>
	+ Note that the Anaconda Python download comes with an IDE (integrated development environment) called Spyder. This is the IDE the class will use. While there are other excellent options to code in Python (such as Jupyter), this class will use Spyder.

## Technical Assistance

If you experience any errors or problems while in your online course, contact 24/7 Canvas Support through the Help link within Canvas.  If you experience computer difficulties, need help downloading a browser or plug-in, or need assistance logging into a course, contact the IS Service Desk for assistance. You can call (541) 737-8787 or visit the [Service Desk](https://oregonstate.teamdynamix.com/TDClient/Requests/TicketRequests/NewForm?ID=Dr9c0T7BaSI_) online.

## Ecampus Reach Out for Success

University students encounter setbacks from time to time. If you encounter difficulties and need assistance, it’s important to reach out. Consider discussing the situation with an instructor or academic advisor. Learn about [resources that assist with wellness and academic success](https://counseling.oregonstate.edu/reach-out-success); you can also access the **Resources** tab in Canvas’ global navigation menu for additional information.

Ecampus students are always encouraged to discuss issues that impact your academic success with the [Ecampus Success Team](https://ecampus.oregonstate.edu/services/student-services/). Email ecampus.success@oregonstate.edu to identify strategies and resources that can support you in your educational goals.

### For mental health:

Learn about [counseling and psychological resources for Ecampus students](https://counseling.oregonstate.edu/main/ecampus-students). If you are in immediate crisis, please call or text the Suicide and Crisis Lifeline at 988 or Crisis Text Line by texting 741-741.

### For financial hardship:

Any student whose academic performance is impacted due to financial stress or the inability to afford groceries, housing, and other necessities for any reason is urged to contact the Director of Care for support (541-737-8748).

## Measurable Student Learning Outcomes

1. **Distinguish and discuss** data science and analytics terminologies, the data science lifecycle, ethics in data science, and data types and formats. (TCO 1)
2. **Build and illustrate** data gathering, cleaning, preparation, visual exploration, and transformation using Python Programming, visual elements, and discussion. (TCO 2)
3. **Distinguish and discuss** unsupervised and supervised machine learning methods including similarities and differences, the nature and format of the data required in different cases, and the overall goals (such as prediction vs. discovery). (TCO 3)
4. **Program** the unsupervised method ofk- means clustering using the Python programming language, including the creation of appropriate visual elements, and discussion and illustration of results. (TCO 4)
5. **Program** the supervised method ofdecision tree modeling using Python programming, appropriate visual elements, discussion and illustration of results, and illustration of prediction accuracies including the use of confusion matrices. (TCO 5)
6. **Create** a final report that applies and utilizes the data science life cycle, data science communication methods, decision science, and data analysis to offer actionable conclusions. (TCO 6)

## Bacc Core / Slash Course / WIC

*This course is not Bacc Core*

*This course is not WIC*

## Evaluation of Student Performance [Note: This may be updated minorly during the dev]

* Discussions – 200 points
* Project-Assignments - 400 points
* Quizzes – 150 points
* Final Class Project – 100 points
* Exercises – 150 points

**Total – 1000 points**

##

## Letter Grade

| Grade | A | A- | B+ | B | B- | C+ | C | C- | D+ | D | D- | F |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Percent Range | *93-**100* | *90-**92* | *87-**89* | *83- 86* | *80- 82* | *77- 79* | *73- 76* | *70- 72* | *67- 69* | *63- 66* | *60- 62* | *0- 59* |

## Course Policies

### Discussion Participation

Students are expected to participate in all graded discussions. While there is great flexibility in online courses, this is not a self-paced course. You will need to participate in discussions on at least two different days each week, with your first post due no later than Wednesday evening, and your second and third posts due by the end of each week.

### Late Work Policy

This course has a variety of deliverables, such as Assignments (which build into the final project), asynchronous discussions, quizzes, exercises, and a final project. Each deliverable has its own due dates. For example, Discussions have two (2) due dates. The first due date is for the main or primary post and the second due date is for peer responses. Some deliverables can be submitted later than the due date for partial credit, while others (such as discussions) cannot be later than the final due date. Please see each deliverable for specific due dates.

### Proctored Exams

There are no proctored examinations in this class.

### Incompletes [ECAMPUS RECOMMENDED]

When a requirement of a course has not been completed for reasons acceptable to the instructor and the rest of the academic work is passing, a report of “I” (incomplete) may be made and additional time granted, according to Academic Regulation 17 of [OSU Academic Regulations](https://catalog.oregonstate.edu/regulations/).

If you are having any difficulty that might prevent you completing the coursework, please don’t wait until the end of the term; let me know right away.

### Statement Regarding Religious Accommodation

Oregon State University is required to provide reasonable accommodations for employee and student sincerely held religious beliefs.  It is incumbent on the student making the request to make the faculty member aware of the request as soon as possible prior to the need for the accommodation. See the [Religious Accommodation Process for Students](http://eoa.oregonstate.edu/religious-accommodation-policy).

**Class Participation and Building Community**
Active interaction with peers and your instructor is essential to everyone’s success in this online course. I encourage you to please practice the following:

* Value the diversity of the class. Recognize and respect the experiences, abilities, and knowledge each person brings to our learning environment.
* Challenge others’ ideas with the intent of facilitating growth. Acknowledge your peers' contributions and highlight areas of further inquiry.
* Be open to being challenged on your ideas or preconceptions.
* Practice self-awareness in your communication with peers and consider that your comments may affect others unintentionally.
* Assume the best of your classmates and instructor and expect the best from them. They will do the same with you.
* Manage your time and maintain a plan throughout the class that allows you enough time to complete required elements by or before due dates.
* If you are a member of a team, be sure to actively communicate and participate.

### Expectations for Student Conduct

Student conduct is governed by the university’s policies, as explained in the Student Conduct Code (<https://beav.es/codeofconduct>). Students are expected to conduct themselves in the course (e.g., on discussion boards, email postings) in compliance with the university's regulations regarding civility.

### Academic Integrity

It is important that you understand what student actions are defined as academic misconduct at Oregon State University.  The OSU Libraries offer a [tutorial on academic misconduct](https://guides.library.oregonstate.edu/c.php?g=286121&p=3896378), and you can also refer to the [OSU Student Code of Conduct](https://beav.es/codeofconduct) and [the Office of Student Conduct and Community Standards](https://studentlife.oregonstate.edu/studentconduct/student-info) for more information.  More importantly, if you are unsure if something will violate our academic integrity policy, ask your professors, GTAs, academic advisors, or academic integrity officers.

Academic misconduct, or violations of academic integrity, can fall into seven broad areas, including but not limited to: cheating; plagiarism; falsification; assisting; tampering; multiple submissions of work; and unauthorized recording and use.

In this course, AI tool use is restricted. Specifically, AI, Google, and other research and search methods can be very productive in learning new concepts. However, if you are asked to write or explain something, be sure that you are doing this yourself so as to gain the experience.

Additional details will be shared in course activity and assignment prompts. I encourage you to contact me if you have questions about the use of AI tools so that, together, we can ensure that we are using these tools in productive and ethical ways.

## TurnItIn

Your instructor may ask you to submit one or more of your writings to Turnitin, a plagiarism prevention service. Your assignment content will be checked for potential plagiarism against Internet sources, academic journal articles, and the papers of other OSU students, for common or borrowed content. Turnitin generates a report that highlights any potentially unoriginal text in your paper. The report may be submitted directly to your instructor or your instructor may elect to have you submit initial drafts through Turnitin, and you will receive the report allowing you the opportunity to make adjustments and ensure that all source material has been properly cited. Papers you submit through Turnitin for this or any class will be added to the OSU Turnitin database and may be checked against other OSU paper submissions. You will retain all rights to your written work. For further information, visit [Academic Integrity for Students: Turnitin – What is it?](http://guides.library.oregonstate.edu/c.php?g=286121&p=1906660)

## Statement Regarding Students with Disabilities

Accommodations for students with disabilities are determined and approved by Disability Access Services (DAS). If you, as a student, believe you are eligible for accommodations but have not obtained approval, please contact DAS immediately at 541-737-4098 or at <http://ds.oregonstate.edu>. DAS notifies students and faculty members of approved academic accommodations and coordinates implementation of those accommodations. While not required, students and faculty members are encouraged to discuss details of the implementation of individual accommodations.

## Accessibility of Course Materials

All materials used in this course are accessible. If you require accommodations please contact [Disability Access Services (DAS)](http://ds.oregonstate.edu/home/).

Additionally, Canvas, the learning management system through which this course is offered, provides a[vendor statement](https://www.canvaslms.com/accessibility)certifying how the platform is accessible to students with disabilities.

## Tutoring and Writing Assistance

# You can connect live with experienced online tutors by accessing TutorMe in the side navigation bar of your Canvas course. You are eligible for up to 5 hours of tutoring each week. To learn more, go to [Online Tutoring - Overview](https://ecampus.oregonstate.edu/services/student-services/online-tutoring/).

To get help with any form of writing, you can contact [Oregon State Online Writing Support](https://writingcenter.oregonstate.edu/ows) for feedback via email or live Zoom appointment.

## Academic Calendar

All students are subject to the registration and refund deadlines as stated in the Academic Calendar: <https://registrar.oregonstate.edu/osu-academic-calendar>.

## Student Bill of Rights

OSU has twelve established student rights. They include due process in all university disciplinary processes, an equal opportunity to learn, and grading in accordance with the course syllabus: <https://asosu.oregonstate.edu/advocacy/rights>.

## Student Learning Experience Survey

During Fall, Winter, and Spring term the online Student Learning Experience surveys open to students the Wednesday of week 9 and close the Sunday before Finals Week. Students will receive notification, instructions, and the link through their ONID email. They may also log into the survey via MyOregonState or directly at <https://beav.es/Student-Learning-Survey>. Survey results are extremely important and are used to help improve courses and the learning experience of future students. Responses are anonymous (unless a student chooses to “sign” their comments, agreeing to relinquish anonymity of written comments) and are not available to instructors until after grades have been posted. The results of scaled questions and signed comments go to both the instructor and their unit head/supervisor. Anonymous (unsigned) comments go to the instructor only.

## Course Content

General Reference for the entire course:

<https://gatesboltonanalytics.com/> (I own these resources)

Books:

(Recommended but not required)

1. <https://jakevdp.github.io/PythonDataScienceHandbook/>
2. <https://scientistcafe.com/ids/IDS.pdf>
3. <https://people.smp.uq.edu.au/DirkKroese/DSML/DSML.pdf>
4. https://covid19.uthm.edu.my/wp-content/uploads/2020/04/Data-Science-from-Scratch-First-Principles-with-Python-by-Joel-Grus-z-lib.org\_.epub\_.pdf

[This is largely set but may have minor updates during the dev]

| **Week/****Module** | **Topics** | **Learning Materials** | **Assignments** |
| --- | --- | --- | --- |
| 1 | Introduction to concepts in Data Science, AI, Machine Learning, the Data Science Lifecycle, Ethics in Data ScienceGetting Python/Anaconda  | https://gatesboltonanalytics.com/Just Python<https://www.w3schools.com/python/default.asp>Getting Anaconda/Python<https://www.anaconda.com/download> | Discussion QuizAssignment |
| 2 | Python and Sklearn for Data Science | Just Python<https://www.w3schools.com/python/default.asp>Sklearn<https://scikit-learn.org/stable/>Using Python for Preparing Data (Python Code Example)<https://gatesboltonanalytics.com/?page_id=351> |  |
| 3 | Data GatheringAPIsPython Packages and Sklearn | API Basics:<https://gatesboltonanalytics.com/?page_id=245>APIs in Python:<https://gatesboltonanalytics.com/?page_id=254>**Open/Free Datasets:**Kaggle Datasets: <https://www.kaggle.com/datasets>Google Datasets<https://datasetsearch.research.google.com/>.gov datasets<https://resources.data.gov/resources/govt-data-hubs/>data.gov<https://data.gov/>Open Datasets<https://datahub.io/collections><http://opendata.cern.ch/>Earth Data<https://www.earthdata.nasa.gov/>World Health Org<https://apps.who.int/gho/data/node.home>Industry Datasets<https://www.bfi.org.uk/industry-data-insights>Tableau Free Public Datasets<https://www.tableau.com/learn/articles/free-public-data-sets>Azure Open Datasets<https://azure.microsoft.com/en-us/products/open-datasets>World Bank Open Data<https://data.worldbank.org/>Data World Open Data<https://data.world/datasets/open-data>  | AssignmentDiscussion |
| 4 | Data Formats and Types | Data Types and Formats:<https://gatesboltonanalytics.com/?page_id=235>Using Python to convert data formats and types<https://gatesboltonanalytics.com/?page_id=351>  | QuizDiscussion or Exercise (TBA) |
| 5 | Data Cleaning and Preparation Part 1Cleaning Record DataExploratory Data Analysis (EDA) | Data Cleaning:<https://gatesboltonanalytics.com/?page_id=403>Preparing Data for Analysis in Python:<https://gatesboltonanalytics.com/?page_id=343>Using Python to convert data formats and types<https://gatesboltonanalytics.com/?page_id=351>  | AssignmentDiscussion |
| 6 | Data Cleaning Part 2:NormalizationTransformationFeature EngineeringLabeled vs unlabeled dataData BalanceHarmonization | Data Cleaning:<https://gatesboltonanalytics.com/?page_id=351><https://www.digitalocean.com/community/tutorials/normalize-data-in-python>Discretization in Python<https://www.youtube.com/watch?v=MtKYXoXcwB8>Creating Training and Testing Data to Train Supervised Models<https://gatesboltonanalytics.com/?page_id=515>  | QuizDiscussion or Exercise (TBA) |
| 7 | Unsupervised Learning - Clustering | Clustering (slides, code, distance metrics):<https://gatesboltonanalytics.com/?page_id=256>More on kmeans clustering:<https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/><https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1> <https://www.w3schools.com/python/python_ml_k-means.asp> Clustering in Python<https://gatesboltonanalytics.com/?page_id=262>KMeans Clustering in Sklearn/Python<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html> | AssignmentDiscussion |
|  |  |  |  |
| 8 | Supervised Learning Methods | Training and Testing Data Preparation:<https://gatesboltonanalytics.com/?page_id=515> | AssignmentDiscussion |
| 9 | Supervised Learning -Decision Tree  | Decision Trees:<https://gatesboltonanalytics.com/?page_id=278>More on Decision Trees:<https://www.w3schools.com/python/python_ml_decision_tree.asp>Decision Trees in Python<https://gatesboltonanalytics.com/?page_id=282><https://scikit-learn.org/stable/modules/tree.html>Decision Tree Classification in Sklearn/Python<https://scikit-learn.org/stable/modules/tree.html#classification>Confusion Matrices<https://www.w3schools.com/python/python_ml_confusion_matrix.asp> | QuizDiscussion or Exercise (TBA) |
| 10 | data science communication methods, communication with a diverse audience, decision science, and how to use data analyses to create actionable conclusions and reports. | Note: I will be building resources here but there are also excellent web resources:<https://towardsdatascience.com/communicating-as-a-data-scientist-why-it-matters-and-how-to-do-it-well-f1c34d28c7c4>Decision Science vs. Data Science<https://www.linkedin.com/pulse/decision-science-vs-data-why-you-should-care-ramon-chen/>Communication in Data Science<https://www.thedataincubator.com/blog/2022/10/13/improve-your-data-science-communication/> | DiscussionFinal Project |
| Finals | Review and Discussion | TBA | Final Project |