# **Syllabus for DATASCI 410 Data Science: Methods for Data Analysis**

|  |  |  |
| --- | --- | --- |
|  | Instructor: Summer Elasady  summe@uw.edu  Office Hours: upon request | Location: Puget Sound Plaza, Room 504  Dates: 01/07/2019– 03/12/2019  Times: Tuesdays 6-9pm |

# Course Description:

This is the second of three courses in the Data Science program. This course provides students with a broad introduction to the theoretical and methodological basis of data science. The focus of the course is on data exploration and visualization, statistical theory, and theory or linear machine learning models. By the end of this course, you will be able to explore complex relationships data, apply statistical methods, and confidently understand the basis of machine learning algorithms.

# Course Learning Objectives:

After successfully completing this course, you will be able to:

* apply the data science process to a business problem including determining data requirements, exploring the data, and presenting actionable results and recommendations;
* explore complex data relationships and present results in an insightful manner to a non-technical audience;
* apply basic concepts of probability and statistics including conditional probability, sampling, and hypothesis testing;
* generalize the theory and practice of linear models as a foundation for machine learning; and
* apply basic time series models for forecasting, simple text analytics, and unstructured data analysis.

# Course Format:

This is a cohort-based, online course, meaning all materials you will need to complete the course are included in the Canvas site, and that you and your classmates will progress through the course together, week by week. You will see demonstrations of the lab exercises and try some of those by yourself. There will be online discussions about the lesson topics and other activities. You will be able to do your exercises, quizzes, and assignments on your own time.

You will use Canvas LMS as the main site to track the weekly tasks and access all the materials you need to complete the exercises and assignments. You will also use Canvas to submit your assignments and take weekly quizzes. Your grades and assignment feedback will be posted in Canvas.

# Course Materials:

There are no required textbooks for this course. However, we have a list of supplementary textbooks and online resources useful for this course:

## Data Science

* [*Data Science for Business*](http://shop.oreilly.com/product/0636920028918.do), Foster Provost and Tom Fewcett, 2013, ISBN-13: 978-1449361327
* [*Data Science from Scratch*](http://shop.oreilly.com/product/0636920033400.do), Joel Grus, 2015, ISBN-13: 978-1491901427
* [*Python Data Science Handbook*](http://shop.oreilly.com/product/0636920034919.do), Jake VanderPlas, 2016, ISBN-13: 978-1491912058

## Probability and Statistics

* [*Statistics Done Wrong, the Woefully Complete Guide*](https://www.statisticsdonewrong.com/), Alex Reinhart, 2015, ISBN-13: 978-1593276201
* [*Naked Statistics, Removing the Dread From Data*](http://books.wwnorton.com/books/Naked-Statistics/), Charles Wheelan, 2014, ISBN-13: 978-0-393-34777-7
* [*Errors, Blunders and Lies*](https://www.crcpress.com/Errors-Blunders-and-Lies-How-to-Tell-the-Difference/Salsburg/p/book/9781498795784), David Salsburg, 2017, ISBN-13: 9781498795784
* [*Statistics in a Nutshell*](http://shop.oreilly.com/product/9780596510497.do), Paul Watters and Sarah Boslaugh, 2009, ISBN-13: 978-1449316822
* [*What is a P-Value Anyway: 34 Stories to help you actually understand statistics*](https://books.google.com/books/about/What_is_a_P_value_Anyway.html?id=KM9FAQAAIAAJ), Andrew Vickers, 2010, ISBN-13: 9780321629302

## Python

* [*Python for Data Analysis*](http://shop.oreilly.com/product/0636920050896.do), 2E, Wes McKinney, 2017, ISBN-13: 978-1491957660

# Technical Requirements:

Your course uses the following technology. Please [check that the Hardware/Software of your device](http://uwodashboard.pce.uw.edu/techTest.asp) meets the requirements.

|  |  |
| --- | --- |
| **Technology** | **Hardware/Software** |
| Canvas LMS | Modern Browser (Chrome, Firefox, or Safari), Bandwidth (Internet Speed), |
| Tutorial Videos with MediaAMP | Bandwidth, Speakers or headphones |
| PDF Viewer | [Adobe Acrobat Reader](https://get.adobe.com/reader/) |
| Word-processor with spell-check | Google Docs or [MS Office](https://itconnect.uw.edu/wares/uware/microsoft/microsoft-office-365-proplus/). |
| Jupyter Notebook with Python 3.\* | [Anaconda Distribution Download](https://www.anaconda.com/download/) |

# Course Topics:

|  |  |  |  |
| --- | --- | --- | --- |
| Part 1—Data Visualization | Part 2—Statistical Analysis | Part 3— Linear Models | Part 4— Other Machine Learning Models |
| **Lesson 1**—Data Exploration Part 1  **Lesson 2**—Data Exploration Part 2 | **Lesson 3**—Combinatorics & Probability Distributions  **Lesson 4**— Sampling & Hypothesis Testing  **Lesson 5**—Introduction to Bayes Theorem | **Lesson 6**—Introduction to Regression  **Lesson 7**—Regression & Regularization  **Lesson 8**—Time Series Analysis | **Lesson 9**—Näive Bayes  **Lesson 10**—Basic Text Analytics |
| **Milestone 1**—Data Visualization Complementary Views | **Milestone 2**—Hypothesis Simulation | **Milestone 3**—Regression Models | **Milestone 4**—Independent Project |

# Student Assessment:

To successfully complete this course, you must:

* Attend and participate actively in class activities
* Answer the quiz questions
* Complete the lesson exercises
* Submit the milestone projects

**You will need to complete and submit all components of each assignment, quiz, and project, and earn an overall average score of 80% or more to pass this course.** Your grade for this course will be recorded on your transcript as SC (satisfactory completion) or USC (unsatisfactory completion).

## Grading Table:

Your grades are based on the following components:

# Late Work Policy:

No late work will be accepted after 1 week from the due date.

|  |  |
| --- | --- |
| **Component** | **Percentage** |
| Attendance and Participation | 16% |
| Quizzes | 20% |
| Lesson Exercises | 20% |
| Milestone Projects 1, 2, and 3 | 24% |
| Milestone Project 4 | 20% |

**Attendance and Participation:** Attendance is critical to the successful completion of this course; you cannot miss more than 2 classes. You are expected to participate in online discussions and activities.

**Quizzes:** The quiz questions are based on the basic application of important concepts from the weekly lessons.

**Lesson Exercises:** Each lesson provides an opportunity for you to practice the skills demonstrated and then apply the skills toward the completion of a milestone project.

**Milestone Projects:** At the end of each part of the course, a milestone project is due representing the application of the skills and knowledge acquired during the course lessons.

# Student Code Policy:

The University o fWashington's Student Conduct Code applies to all students, including students enrolled in UW Professional & Continuing Education courses. Students are expected to maintain the highest standards of academic responsibility. Plagiarism and other kinds of academic misconduct are considered serious offenses at the UW. Plagiarism is using someone else’s words or ideas without proper citation. It can range from failure to credit a single sentence or paragraph to passing off an entire article, speech or another student’s paper as one’s own. Instances of academic dishonesty for noncredit courses are handled by the University of Washington Professional & Continuing Education Committee on Academic Conduct. If evidence of academic misconduct is established, the student will be given a failing grade for the course and any request for a refund of course or other fees will be denied.

# Disability Services:

The Disability Services Office strives to help make the UW community more accessible for all. If you are a non-degree student seeking accommodation for a permanent or temporary disability, contact the office for more information and assistance. You can reach Disability Services at 206-543-6450 or [dso@uw.edu](mailto:dso@uw.edu).

# About the Instructor

SUMMER RAE ELASADY

Summer Elasady is a bioinformatic scientist who’s been working in the software industry for more than a decade. She sold her first database-driven software company at the age of 26 and was then hired to manage the fraud detection team for GoDaddy.com. She’s spent the past three years using Python, PySpark, R and SQL to improve alignment algorithms for genome/transcriptome assembly and analysis.

Prior to working with UW Professional & Continuing Education, Elasady designed training courses for professional software companies including GoDaddy.com and MyProcess and taught undergraduate courses at the University of Oregon. She’s a member of the Puget Sound Python Programming group and has helped designed STEM courses for Washington high schools as a fellow for the Washington Alliance for a Better School System. She’s passionate about teaching Python and making the STEM fields more accessible. Elasady earned a Teaching Excellence Award in both 2017 and 2018 for her work with certificate programs.