SYLLABUS FOR
DATASCI 410 DATA SCIENCE: METHODS FOR DATA ANALYSIS

Mohamed Mneimneh
Instructor

Datasc2@uw.edu
Office Hours: upon request

Sahar El Turk
Instructional Assistant

Datasc2@uw.edu

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.
- Apply basic time series models for forecasting, simple text analytics, and unstructured data analysis

COURSE FORMAT:

This is a self-paced online course, meaning you will be responsible for keeping progress and managing your time in order to complete the course within 4 months. You will see demonstrations of the lab exercises and try some of those by yourself. You will be able to do your exercises, quizzes, and homework assignments on your own time.

UW Start Learning LMS is your course site, you will use it to access your lessons, get all the materials you need to complete the exercises and assignments, take quizzes, submit your homework, communicate with other students and instructional team, and check your progress. Your grades and homework assignment feedback will be posted in the LMS.

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*, Foster Provost and Tom Fewcett, 2013, ISBN-13: 978-1449361327
- *Data Science from Scratch*, Joel Grus, 2015, ISBN-13: 978-1491901427
- *Python Data Science Handbook*, Jake VanderPlas, 2016, ISBN-13: 978-1491912058

Probability and Statistics

- *Statistics Done Wrong, the Woefully Complete Guide*, Alex Reinhart, 2015, ISBN-13: 978-1593276201
- *Naked Statistics, Removing the Dread From Data*, Charles Wheelan, 2014, ISBN-13: 978-0-393-34777-7
- *Errors, Blunders and Lies*, David Salsburg, 2017, ISBN-13: 9781498795784
- *Statistics in a Nutshell*, Paul Watters and Sarah Boslaugh, 2009, ISBN-13: 978-1449316822
- *What is a P-Value Anyway: 34 Stories to help you actually understand statistics*, Andrew Vickers, 2010, ISBN-13: 9780321629302

Python

- *Python for Data Analysis*, 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 meets the requirements.

Technology	Hardware/Software
UW Start Learning LMS	Modern Browser (Chrome, Firefox, or Safari), Bandwidth (Internet Speed),
Tutorial Videos with MediaAMP	Bandwidth, Speakers or headphones
PDF Viewer	Adobe Acrobat Reader
Word-processor with spell-check	Google Docs or MS Office.
Jupyter Notebook with Python 3.*	Anaconda Distribution 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:

- Participate actively in class discussion activities
- Answer the quiz questions
- Complete the lesson assignments
- 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:

Component	Percentage
Active Participation	18%
Quizzes	25%
Lesson Assignments	27%
Milestone Projects	30%

LATE WORK POLICY:

All course work is due within 4 months of starting this course. No assignments are accepted after your deadline.

Active Participation: Participation is measured through focused discussion activities which present a question and require you to explain your stance on the answer. You are expected to participate in focused discussion activities.

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

Lesson Exercises: Each lesson provides the 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 of Washington'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.

About the Instructional team

Mohamed Mneimneh has been with SAP for more than nine years, holding roles as research scientist, software engineer, product manager and principal solution adviser. Mneimneh helps companies digitally transform their business using machine learning and data science. He holds a Ph.D. in computer and electrical engineering with a focus on machine learning from Marquette University.

Sahar El Turk holds a B.S in Computer Science, Masters in Business Administration and an EdD in Interdisciplinary Leadership from Creighton University. Sahar is a trilingual person who has a passion for online education. She learned Data Science while pursuing her Doctorate degree and she used the mixed methodology to conduct her Doctoral dissertation which was about the perceived online education barriers through the voices of administrators and faculty and where she used the exploratory factor analysis to identify the factors of the perceived barriers to online education.