CSE 252 – Data Science

Subject Code and Course Number: CSE 252
Course Title: Data Science
Number of Credits:
Instructor Name: *Habet Madoyan*
Email Address: hmadoyan@aua.edu
TA: Karen Mkhitaryan karen_mkhitaryan@edu.aua.am
Office Location: 317W
Office Hours: Stop by anytime
Term/Year: Summer 2018
Class Schedule: Tuesday & Thursday 09:30-12:20 with 10 minutes break
Prerequisites: *CS 121- Data Structures*
Co-Requisites:

Course Description: This course aims to introduce students to the world of data science. Students will gain the skills that are transforming entire industries from healthcare to internet marketing and beyond. In this course, students will gain a hands-on introduction to using R programming language for reproducible data analysis. Students will define the data science process, including data acquisition, data munging, exploratory data analysis, visualization and modeling real world data. The course will include using R and R packages tools for analysis of both structured and unstructured data sources, as well as writing reproducible data analysis reports with R Markdown and creating personalized interactive graphics applications.
Additional Information About the Course: No prior knowledge in programming or data science is required. Students will be required to complete individual projects in order to develop their ability to explore and visualize data on their own. Three hours of instructor-led class time, including discussions and problem solving, plus six hours of homework per week.

Reading Materials: *These materials will help you better navigate through the course (these are not mandatory readings):*
"Introduction to R" by O'Reilly;
"Developing Data Products in R" by Brian Caffo,
"Report Writing for Data Science in R" by Roger Peng,
"Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining" by Simon Munzert, Christian Rubba, Peter Meißner, and Dominic Nyhuis, Wiley, 2015

Software:
R https:/cran.r-project.org (freeware)
R Studio, https://www.rstudio.com (freeware)

Schedule & Topics*: Tentative list of topics*

	Topics	Assignment
Week 1	Data Science Basics/R language basic/ Subsetting data in R/ Principles of Tidy data, R markdown	HW1
Week 2	Data Cleaning, Data Transformation and Data Visualization	HW2
Week 3	Regular Expressions, working with text,	HW3
Week 4	Working with API's and Web scrapping with R	HW4 (Group homework)
Week 5	Midterm/Building data products: R Shiny	HW5

Week 6: The classes will be on Monday/Tuesday for this week	Discussing projects/Spatial Analysis in R, Note	HW6
Week 7	No class	
Week 8	Project discussions and project presentations	

Student Learning Outcomes:
The following chart shows alignment between course-specific and program student learning outcomes and program goals.

Course-based Student Learning Outcomes *In this course, students will be able to:*	*Program Student Learning Outcomes* *Students will be able to:*	*Program Goal*
• Utilize R and R Studio to acquire, clean, and manipulate data, and to perform data analysis and model fitting using R • Create interactive visualizations using R and R packages • Learn how to scrape and analyze data off the web • Analyze text using R Studio text mining packages • Create reproducible data analysis reports with R Markdown	1.1. Use concepts and methods of mathematical disciplines relevant to mathematical modeling 2.3 Formulate and critically assess problems and sub tasks including conduct of appropriate research, identification of sources and investigative techniques 3.2 Develop and implement software applications in one or more programming languages	1. Equip students with knowledge and advanced skills in mathematical reasoning, problem solving, modeling and scientific computation 2. Train students for careers and advanced studies in a wide range of applied mathematical and computational disciplines 3. Prepare students for development of scientific, engineering and industrial software applications

Method of Evaluation

Student learning will be evaluated on the basis of the following weighted components:

	Points	Percentage
Homework	200	20%
Midterm	350	35%
Final Project	300	30%
Attendance	50	5%
Datacamp	100	10%

Datacamp Assignment:
You will receive free access to all datacamp.com courses. Datacamp is a leading provider for data science online courses. To complete this assignment, you need to collect at least 20,000 EXP which translates to around 4 courses. As python is another programming language widely used for data

science, only EXP gained from Python tracks will be counted for this assignment. Python courses are listed here. Send me the email that you want to use for Datacamp ASAP.

Attendance:
The student is allowed to miss 3 classes (excused and unexcused) during the whole course. If the student was absent more than 3 times during the course, he will lose 5% of the grade. If he has absented less than 4 times, the whole 5% of the grade will be given.
The attendance check will be done during the class at the random period of time. Student gets "present" if he was at the class during the attendance check.

Mid-Term Exam:
The purpose of the mid-term exam is to assess students' progress in learning how to use R language to manipulate and visualize various types of data structures and samples. Students will receive feedback through the in-class mid-term review as well as written feedback on their mid-term exams. Students are encouraged to attend office hours for more individualized guidance.

Group Project
You need to form a group at the end of the second week. More details on the final project will be given during the course. Please note that Homework 4 is a group homework.

Late Policy
The late assignments will be accepted on instructor's discretion, only if the student submits convincing evidence of a medical or other emergency that made completing the assignment at the scheduled time impossible.

Make-up Procedures
Make-up assignment, exam, and quiz will be given at the instructor's discretion. Students must submit convincing evidence of a medical or other emergency that makes completing an assignment or taking an exam or quiz at the scheduled time impossible.

Policy on Grade Appeal
Students are entitled to appeal grades in line with the university's Grades Policies policy which is available online at http://policies.aua.am

Standards for Academic Integrity
Students are required to conduct themselves in an academically responsible and ethical manner in line with the Student Code of Ethics. Acts of academic dishonesty impair the academic integrity of AUA and create an unfair academic advantage for the student involved and other member(s) of the academic community. These acts are subject to disciplinary measures as prescribed in the AUA Code of Student Ethics http://policies.aua.am/policy/10

The Student Code of Conduct can be found at http://policies.aua.am/policy/101

Special Needs:
Students requiring special accommodations for learning should contact the Center for Student Success by the end of the Drop/Add period with such requests. studentsuccess@aua.am, http://studentsuccess.aua.am/disability-support-services/