General Assembly Data Science SF 10

Logistics:

- **Dates:** 10/1-12/17, MW 6:30-9:30
- There will be one all day Saturday session to make up 2 class sessions, tentatively scheduled for 11/8
- **Location**:
  - 580 Howard Street (GA Lofts) till mid November
  - 225 Bush Street after mid November
- **Instructor:** Francesco Mosconi – f@mosconi.me
- **Experts-in-Residence (TA's): •**
  - Craig Sakuma - craig.sakuma@gmail.com
  - Michael Keba - mjkeba@gmail.com
- **Office Hours:** Schedule weekly with Craig and Michael

COURSE DESCRIPTION

This course is a practical approach to the knowledge and skills required to excel in the field of data science. Through various case studies, real-world examples and guest speakers, students will be exposed to the basics of data science, fundamental modeling techniques, and various other tools to make predictions and decisions about data. Students will gain practical computational experience by running machine learning algorithms and learning how to choose the best and most representative data models to make predictions. Students will be using Python throughout this course.

COURSE MATERIALS

Students are required to bring a laptop to class everyday. Please come to the first class with Continuum Anaconda (http://continuum.io/downloads) installed, as detailed in the Pre-Work document.

COMPLETION REQUIREMENTS

In order to receive a General Assembly Certificate in Data Science, upon completion of the course, students must:

- Complete and submit 80% of all course assignments (homework, labs, quizzes). Students will receive feedback from instructors on their assignments on a timely basis. Students who miss more than 20% of assignments will not be eligible for the course certificate.
- Complete and submit the course final project, earning a satisfactory grade by completing all functional and technical requirements on the project rubric, including delivering a presentation.

Assignments, milestones and feedback throughout the course are designed to prepare students to deliver a quality course project.

COURSE OUTLINE

The weekly schedules for lecture content, lab content, and homework assignments are subject to change according to the needs & desires of the class. Dates for guest speakers may change according to speaker availability.

Course Overview

1. Basics of Data Science
2. Machine Learning
3. Advanced techniques
4. Guest speakers/special topics
5. Final project

Calendar

Lesson	Date	Day	Topic	HW	FP
1 U1-1	1-Oct	W	Introduction to Data Science		
2 U1-2	6-Oct	M	Getting data		
3 U1-3	8-Oct	W	Cleaning and exploring data	1	Title due
	~~13-Oct~~	~~M~~			
4 U2-1	15-Oct	W	Intro to Machine learning	2	
5 U2-2	20-Oct	M	Regression and regularization		
6 U2-3	22-Oct	W	Dimensionality reduction	3	Data sources
7 U2-4	27-Oct	M	Dimensionality reduction, Recap		
8 U2-5	29-Oct	W	K-means, Clustering	4	
9 U2-6	3-Nov	M	Logistic regression		
10 U2-7	5-Nov	W	Database technologies, Unstructured data, map reduce	5	Elevator pitch
11 U3-1	8-Nov	S	Database technologies, Unstructured data, map reduce, Recommendation systems		
	~~10-Nov~~	~~M~~			
12 U3-2	12-Nov	W	Decision trees, RF	Midterm	
13 U3-3	17-Nov	M	Non linear models, SVM		
14 U3-4	19-Nov	W	Text Mining	7	Project proposal
15 U3-5	24-Nov	M	Naïve bayes		
	~~26-Nov~~	~~W~~		8	
16 U3-6	1-Dec	M	Parallel computing		
17 U4-1	3-Dec	W	Guest Speaker	9	Github repo
18 U4-2	8-Dec	M	Guest Speaker		
19 U5-1	10-Dec	W	Final project working session		
20 U5-2	15-Dec	M	Final project presentations		
22 U5-3	17-Dec	W	Final project presentations		