USC Viterbi
School of Engineering

# DATA SCIENCE PROGRAM

## DSCI 560: Data Science Professional Practicum

**(4 units)**

***Syllabus***

**Term-Day-Time**

Fall 2021- Friday 1:00-4:50pm

**Instructor:** Dr. Anna Farzindar
Email: farzinda@usc.edu

### Professor's Office Hours:

**Office:** Physical (GER 202B) or virtual address

**Office Hours:** Friday Before class

Students are advised to make appointments with the professor ahead of time in any event and be specific with the subject matter to be discussed. Students should also be prepared for their appointment by bringing all applicable materials and information.

### Catalogue Description:

Student teams working on external customer data analytic challenges; project/presentation based; real client data, and implementable solutions for delivery to actual stakeholders; capstone to degree.

### Expanded Course Description:

One of the fundamental principles of data science is the ability to live in the data. It is necessary to gain a level of immersion in the information environment to truly apply the diverse skill sets necessary to both become an effective analyst, and provide customers solutions to hard problems. While students will get a sense of this paradigm in many of the courses focused on knowledge and skill enhancement throughout their degree matriculation, the goal of this course is to combine previously learned capabilities and apply them against actual data sets, in real data environment, and toward solving difficult challenges for an external stakeholder.

The Data Science Professional Practicum is a capstone experience designed to allow students exposure to the world of data analytics from the perspective of the organization. Students will work with external stakeholders on a project that answers an organizational problem. Each semester, the data science program will partner with an external organization, which will pose a broad business requirement to the class. Students will team and propose projects around this requirement and be responsible for delivering a data science-based solution to the stakeholder. This product will consist of a not less than 50-page report detailing the proposal, methods utilized in the analysis, the output from the analysis, and

conclusions and recommendations based on the study. Students will also be responsible for a not less than 30-minute presentation on their project to the stakeholder.

A possible example of a capstone project would be: Google volunteers to be the external partner. They obviously are one of the main collectors and brokers of information in the world. One of their areas of philanthropic interest is disease monitoring. Google might pose the question "How can the data Google collects in general could be better utilized to inform on disease, or help mitigate disease, in various areas in the world?" It would be up to the student teams to proposal a project plan, design a study, implement the analysis, and report conclusions and recommendations.

The Professional Practicum (DSCI 560) provides an opportunity to face real-world challenges and to develop your personal innovative project in cooperation with the members of your team. The combination of industry and academic experiences will provide a peek at what to expect as a data science experts and data scientist, in addition to providing project management techniques such as lean six sigma aimed at efficiency and effectiveness.

## Learning Objectives

- Expected learning outcomes for students are:
- To understand how to apply the various engineering and business principles
- studied in Data Science curriculum toward solving an organizational
- challenge
- To handle difficulties associated with defining and organizing a realistic problem
- Statement
- To manage impediments in obtaining information and approval
- To present and sell ideas to higher-level management
- To understand the importance of the need for a continuous exchange between engineers, management and employees in solving an existing problem, given a set of constraints
- To meet aggressive deadlines in a multidisciplinary team effort
- To improve project-based presentation skills, both in-class and in company settings
- To understand the requirements and objectives of customers, how these vary, and how one must tailor a solution to the expectations of a customer
- To understand how data science are viewed in diverse domains; and how solutions and approaches will differ in various fields
- To understand how to work with individuals from diverse domain (engineering, business, etc.) to accomplish a common goal

Students are expected to have foundational knowledge in data management, machine learning, data mining, and data visualization, as well as other topics associated with the Data Science degree program. The course can be taken by students in the ISE Analytics program and/or the Marshall Business Analytics program (being created) as well as the Data Science program. Course is usually taken after other courses in the program. This class will be primarily group work, with some assigned readings, and a major project and presentation that will count for the primary grade in the course.

## Assignments/Reports/Project

The major deliverable from this course is a written project report that reflects a proposed solution, and subsequent results, for an external stakeholder. At the first class meeting, the external partner will be introduced, and a problem/requirement posed. Students will then divide into teams. All projects are team projects. All topics will be unique, and there should be high levels of collaboration on the projects with team members and the professor. Students are required to meet with the professor not less than three times during the semester. Students are required to submit three progress reports throughout the course. There will be a midterm.

There will be three inputs to students' final grades: 1) instructor evaluation; 2) peer review; and 3) evaluation by external stakeholder.

NO PROJECTS WILL BE ACCEPTED LATE, AND STUDENTS MUST PRESENT THEIR FINDINGS ON THE DAY SELECTED DURING THE SEMESTER.

## Class Communication:

USC D2L, Piazza and Blackboard at USC will be used for class communication.

## Grading Schema:

| | |
|---|---|
| Course Project Report and Presentation (Includes peer reviews): | 25% |
| Sponsor Review of Report, Presentation, and Findings: | 20% |
| Assignments: | 15% |
| Quizzes: | 15% |
| Midterm: | 15% |
| Monthly Reports and Instructor Meetings: | 10% |
| **Total** | **100%** |

Grades will range from A through F. The following is the breakdown for grading:

| | |
|---|---|
| 94 - 100 = A | 74 - 76 = C |
| 90 – 93 = A- | 70 - 73 = C- |
| 87 – 89 = B+ | 67 - 69 = D+ |
| 84 – 86 = B | 64 - 66 = D |
| 80 – 83 = B- | 60 - 63 = D- |
| 77 – 79 =C+ | Below 60 is an F |

## Class Structure & Schedule:

Class sequence, dates, topics and guest speakers are subject to change as the semester proceeds. Any revisions will be noted and announced in class in advance.

| Week | Topic | Readings/Notes | Homework |
|---|---|---|---|
| **Week 1** | Orientation, Requirement Outline, Developing Project Ideas, Fundamentals of Project Management | Overview of project management in Data Science<br>Post-class, teams will be formed | Ref: Fundamentals of Project Management (Worksmart) : Joseph Heagney |
| **Week 2** | External Partner presentation, Lean Thinking, Lean Six sigma, Deliverables, DMAIC[1], Analytics in the "real world" | The external sponsor will present and challenge the class<br><br>Proposals briefs (1-3 pages outlining "big idea") will be completed | Team Formation, Project Selection, Proposal Briefs |
| **Week 3** | Value stream mapping , Quality, Variability, and Role of the Analyst | Project Initiation; Prepare Monthly Report<br>Proposal Brief Review – schedule time with instructor | Proposal Briefs Due (assignment 1) |
| **Week 4** | Root Cause Analysis, Teams, People, Lean Engineering | Schedule Time to Meet with Instructor | Process Mapping and Value Stream Mapping (assignment 2) |
| **Week 5** | Understanding the Analytics Product in Business, Improve | You must meet with your instructor this week to review monthly report | NA |
| **Week 6** | Total project management and Theory of Constraints | Project Work Time | First Monthly Report Due (assignment 3) |
| **Week 7** | Mid-Term Review | Half of class dedicated to mid-term review<br>Remainder of class dedicated to project presentation | Team Presentation on Monthly report |
| **Week 8** | Mid-Term, Project Work Time | Mid-Term<br>Schedule meeting with instructor | Second Monthly Report |
| **Week 9** | Stakeholder Session | Stakeholder will be in class to review progress, answer questions, Students will provide the presentations.<br>You must have had your second instructor review meeting by the end of this week | Instructor Meeting |

[1] DMAIC refers to a data-driven quality strategy for improving processes, and is an integral part of the company's Six Sigma Quality Initiative. DMAIC is an acronym for five interconnected phases: Define, Measure, Analyze, Improve, and Control.

| | | | |
|---|---|---|---|
| **Week 10** | Leadership & Innovation in Start-ups | Guest Lecturer on the topics of Data analytics<br>Management, Leadership and entrepreneurship.<br>Correction and discussion of Midterm exam. | |
| **Week 11** | Understanding the Analytics Product in industry | Review expectations of the analyst in the Retail Industry and marketing.<br>Schedule time to meet with instructor | Second Monthly Report Due |
| **Week 12** | Business Data Science | Guest Lecture on the topics of analytics and "Big Data".<br>Schedule time to meet with instructor | |
| **Week 13** | Dress Rehearsal | For those who desire, they will be able to go over their reports and presentations with the instructor, and receive feedback<br>before presenting to the customer | Presentations |
| **Week 14** | Project Presentations | Presentations will conclude<br>All projects must be submitted before final class | Presentations Continue |
| **Week 15** | Projects Due | Team presentation | Final Project Due |
| **Final evaluation** | Stakeholder Summit | Stakeholder will be in class for Student presentations and final reports | Final Project close out |

## Books and Readings:

All books, papers or reports will be available to students in one of three ways: 1) in the USC bookstore; 2) via a CD that the instructor will provide at the beginning of class; and/or 3) via the web.

J. Heagney "Fundamentals of Project Management (Worksmart)" Fourth Edition Edition, 2011.

S. Shaffie, S. Shahbazi, "The McGraw-Hill 36-Hour Course: Lean Six Sigma (McGraw-Hill 36-Hour Courses) Paperback – April 30, 2012.

C. E. Knoeppel, "Installing Efficiency Methods," The Engineering Magazine Company, republished by Google Books, 1917.

***Assigned Reading and Media List:***

A. Farzindar and D. Inkpen "Natural language processing for social media 3rd. Edition" second edition, Morgan & Claypool Publishers, 2020.

M. George, J. Maxey, D. Rowlands, M. Price "The Lean Six Sigma Pocket Toolbook: A Quick Reference Guide to 100 Tools for Improving Quality and Speed and Complexity. January 1, 2004

J. Lean Six Sigma for Supply Chain Management, Second Edition: The 10-Step Solution Process 2nd Edition

McKinsey Global Institute, "The challenge – and opportunity – of 'big data,'" in McKinsey Quarterly, May 2011.

Richard T. Kouzes, Gordon A. Anderson, Stephen T. Elbert, Ian Gorton, and Deborah K. Gracio, "The Changing Paradigm of Data-Intensive Computing," in IEEE Computer, January 2009.

David Bollier, "The Promise and Peril of Big Data," through The Aspen Institute Communications and Society Program, 2010.

James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, "Big data: The next frontier for innovation, competition, and productivity,'" published through McKinsey Global Institute, May 2011.

Viktor Mayer-Schonberger, Kenneth Cukier, "Big Data: A revolution that will transform how we live, work and think," Emaon Dolan/Houghton Mifflin Harcourt, March 2013.

Eric Siegel, "Predictive Analytics: The power to predict who will click, buy, lie, or die," Wiley, Feburary 2013.

Lev Manovich, "Trending: the Promise and the Challenges of Big Social Data," Debates in the Digital Humanities, July 2011.