

Introduction to Data Sciences and Analytics

- Course** UC Berkeley Extension COMPSCI X407.9-007; 2 semester units
Course contents © Stephan Sorger; 2017 Spring #1
- Instructor** Stephan Sorger
Phone: 650.455.4411
email: sorger@berkeley.edu; www.StephanSorger.com
- Meetings** January 30 – March 27, 2017; Monday 6:15 PM – 10:00 PM; San Francisco
Some class meeting dates subject to change

Course Description

This course is designed to provide insights into the concepts of Data Science and Analytics. It provides an overview of various tools such as advanced Microsoft Excel, SQL, and R to analyze huge sets of data and explains the theory of formulating statistical models using regression analysis. Also, the course introduces segmentation theory and methods such as agglomerative clustering using Ward's method.

Course Goals and Learning Objectives

On successful completion, participants will be able to:

- Understand basic statistical concepts
- Execute advanced Microsoft Excel skills for data analysis
- Interpret data and charts from different sources
- Develop linear regression models
- Segment different sets of data
- Understand R and SQL coding

Intended Audience

This course is ideal for managers, junior analysts, engineers or anyone who wants to develop the skills to comprehend huge sets of data to make informed business decisions.

Prerequisites

- Knowledge of basic mathematics
- Basic proficiency with Microsoft Excel

Instructional Methodology

- Lectures on vital areas of data science and analytics
- Case studies of data science concepts applied toward practical problems
- Videos highlighting areas of data science
- Assignments to practice the application of data science principles
- Exams to test marketing analytics concepts and terminology

Reading

Required:

Sorger, Stephan. "Marketing Analytics: Strategic Models and Metrics." Admiral Press/ CreateSpace, 2013. ISBN # 978-1481900300. Print book and Kindle ebook versions available. Buy through Amazon.com.

Website

Go to StephanSorger.com for course content. Request password from instructor.

Grading and Course Components

Grading is calculated from the components shown below, using standard grading cutoff points:
100 – 94 = A, 93 – 90 = A-, 89 – 87 = B+, 86 – 84 = B, 83 – 75 = B-, 74 – 60 = C

	<u>Percent</u>
Assignments (2 x 15 pts. each)	30%
Midterm Exam:	35%
Final Exam:	<u>35%</u>
Total	100%

Assignments

Assignments will be graded using the following grading criteria.

Each assignment will be worth a total of 15 points. Two assignments total to 30 points.

Please do not submit copies of instructor's work (except for data sets) or other students' work.

Please submit Microsoft Word or Adobe PDF files. Please do not submit Excel files.

+ 5 points: Correct Answers

Grading considers correct answers, such as showing work, calculating the correct answers, including screenshots of Microsoft Excel and R sessions, and other elements indicating care taken to the answer process, such as comments in R code.

+5 points: Commentary

Grading considers strong commentary, with good data-based analysis, comparison of results with industry averages and trends, recommendations, external research citations, and critical thinking. This section is vital, not just for the class, but for industry as well.

+5 points: Logistics

Grading considers original work, correct formatting, and delivering the documents by the specified deadline.

Assignments a day+ late will receive a few points off; Assignments more than four days late are not accepted.

Sample Commentary

"The calculations show an average price for identity monitoring services of \$24.67 per month, which is a bit higher than I expected, considering that the data set ranged from \$3.00 to \$35.00. The average price compared well to that reported in a recent Wall Street Journal article, "Identity Monitoring Services: Worth the Money?" (1) (1) Sampson, Joseph. "Identify Monitoring Services: Worth the Money?" Wall Street Journal. April 1, 2016.

Link: www.wsj.com/identity-monitoring-systems-are-they-worth-the-money

Exams

The midterm is closed-book, and tests the following units: 1, 2, 3, 4

The final exam is closed-book, and tests the following units: 5, 6, 7, 8

Academic Ethics Honor Code

All members of the UC Berkeley Extension community are expected to act with honesty, integrity, and respect for others. For further information, please refer to:

Tips for maintaining academic integrity: http://extension.berkeley.edu/upload/academic_integrity.pdf

UC Berkeley Extension Code of Student Conduct: <http://extension.berkeley.edu/upload/studentconduct.pdf>

Professional Behavior

The course does not tolerate any whining, complaining, or other non-professional behavior. We assume you are here to learn. Learning is a partnership between the instructor and the student. Both must engage with a positive attitude to succeed. Students who do not take responsibility for their own success and instead blame others (including the instructor) for their failures, such as low scores on exams and assignments, or poor team performance will be considered "disruptive students" due to the toxic environment they create for others. Disruptive students could be asked to leave the course.

Disabled Student Services

If you require academic accommodations for this course, you should obtain approval from Extension Disabled Student Services. Please contact them at dss-unex@berkeley.edu or (510) 643-5732. If you already have a letter of accommodation from Extension Disabled Student Services, please make an appointment with your instructor to have a confidential discussion of what you will require for this course.

Schedule

Readings are shown in Sorger analytics textbook as “Sorger Ch. 1”, etc.)

January 30, 2017 Meeting 1: Spreadsheets 1

- Administration Review syllabus
- Unit 1 Introduction to Data Sciences (Sorger, Ch. 1, Ch. 6)
- Unit 2 Microsoft Excel: Essentials (Sorger, Ch. 12)
Charts; Copy/Paste; Filter and Sort; Find/Replace; Formatting; Vlookup
- Video Presenting Data in Charts (2:10)
- Case Study In-class case study: Precious Metals; Olympic Medals

February 6, 2017 Meeting 2: Spreadsheets 2

- Unit 3 Microsoft Excel: Tools (Sorger, Ch. 10)
Basic Statistics; Pivot Tables; Solver; Analysis ToolPak
- Video Pivot Tables (7:38)
- Case Study In-class case study: Pivot Tables
- Assignment 1 Due on meeting 3: Krrazy Apps

February 13, 2017 Meeting 3: Spreadsheets 3

- Unit 4 Microsoft Excel: Regression (Sorger, Ch. 3)
Linear regression; R-Squared; Performance tests; F and T tests; P tests; ROC curves
- Video Linear Regression: Statistical Inference (7:29)
- Case Study In-class case study: Apartment; Multivariate
- Assignment 2* Due on meeting 4: Cholera

February 20, 2017 Class Canceled; Federal Holiday: President’s Day

February 27, 2017 Meeting 4: Spreadsheets 4

- Unit 5 Microsoft Excel: Forecasting (Sorger, Ch. 6)
Forecasting; Time series; Causal analysis
- Video Forecasting Inflections (14:21); Backlash (4:31)
- Case Study In-class case study: Acme Real Estate (Part 1)
- Assignment 3* Due on meeting 5: GoPro Forecast

March 6, 2017 Meeting 5: SQL

- Unit 6 SQL Overview and coding
- Video Introduction to MySQL (3:07); MySQL for Excel (4:41)
- Case Study In-class case study: Acme Music
- Midterm Exam Midterm Examination

March 13, 2017 Meeting 6: R and Statistical Programming Languages

- Unit 7 R Module 1: Essentials
Working in R; R functions; Packages; Loading data; Regression analysis
- Video Getting Started with R (16:30)
- Case Study In-class case study: Acme Real Estate (Part 2)
- Assignment 4 Due on meeting 7: Buster Posey

March 20, 2017 Meeting 7: R and Segmentation

- Unit 8 R Module 2: Segmentation using R (Sorger, Ch. 3)
Segmentation; Cluster analysis; Naïve Bayes; Ward’s; K-means; Applications in R
- Video Malcolm Gladwell: Segmentation and spaghetti sauce (17:33)
- Case Study In-class case study: Acme Dog Nutrition

March 27, 2017 Meeting 8: Close

- Tableau Data visualization using Tableau
- Video Introduction to Tableau Public (9:50)
- Final Exam Final examination

* **Not for grade**