

# Fall 2021 Syllabus for 95-851 / 95-451: Making Products Count: Data Science for Product Managers

Fall 2021 Mini-course 1 (6 units): MW 4:40 PM – 6:00 PM EST; Hamburg Hall 1005.

Professor: David Steier ([steier@andrew.cmu.edu](mailto:steier@andrew.cmu.edu))

- Office hours: Immediately after class and Fridays by appointment (send an email to schedule)

Teaching Assistants:

- Riyaz Panjwani ([rpanjwan@andrew.cmu.edu](mailto:rpanjwan@andrew.cmu.edu))
- Yuehan Xiao ([yuehanx@andrew.cmu.edu](mailto:yuehanx@andrew.cmu.edu))

## Course Summary

Product managers engage in a variety of complex activities critical to product success including

- Gathering product requirements
- Prioritizing features
- Forecasting customer demand
- Customer segmentation
- Product pricing
- Identifying buying patterns
- Analyzing and responding to customer feedback

Historically decisions in these areas have often relied on intuition and guesswork, leading to misjudgment of the market and other key factors, and ultimately, product failures. Developments in data science, combining the increasing availability of data from internal and external sources with new algorithms that exploit that data at scale, offer new possibilities for putting product management decisions on a more quantitative and rigorous footing. Students in this course will be introduced to a variety of data science techniques applicable to activities to which product managers typically contribute. These techniques include clustering, classification, A/B testing, and analytics for unstructured data, including clickstreams, text, speech, and images.

This course is for students who are looking for an introduction to applying data science to product management. Backgrounds in basic statistics and some programming experience are required. Students may wish to review the fundamentals of statistics and probability in the free online learning class at <https://oli.cmu.edu/courses/probability-statistics-open-free/>. Hands-on exercises in Python will illustrate the concepts, but please note this is not a Python class. Students who are unfamiliar with Python will get access to online tutorials in DataCamp to build up their Python skills. In-class exercises and weekly assignments will mainly focus on data science techniques and their application to decision-making at various stages of the product life cycle. In the final project, students will select from a variety of data sets to address a product management issue in more depth, from framing the problem through modeling to communicating results.

## Outcomes

The main learning objectives of the course are to enable students to:

1. Identify decision points during the product life cycle where data science techniques are applicable
2. Select from a broad set of metrics, product instrumentation, data sources, modeling and data visualization techniques for use in product management decision-making
3. Apply selected modeling (e.g. classification, clustering, and text analytics) and visualization techniques to product management.
4. Plan and execute a data science project at realistic scale to inform at least one product management decision demonstrating mastery of objectives #1 - #3.

## Instructor: David Steier (PhD, CMU SCS '89)

David Steier joined the CMU faculty in 2018 as a Distinguished Service Professor in the Heinz College School of Information Systems and Management. He teaches courses on data science for product management, managing analytics projects, designing smart systems and artificial intelligence. Prior to joining CMU, David was Managing Director in Deloitte Consulting's Data Science practice. At Deloitte, David helped clients use advanced data analytics and visualization in a variety of industries including health care, banking, retail, manufacturing, telecommunications, media and the public sector. Prior to Deloitte, David was Director in the Center for Advanced Research at PwC, Senior Director of Technology and Business Development at Kanisa, and Managing Director at Scient. In addition to his CMU affiliation, David is also a Lecturer at the University of California Berkeley's School of Information, where he is Course Lead and co-instructor for the data science capstone class in the Masters in Information and Data Science program. David holds a Ph.D. in computer science from Carnegie Mellon and a bachelor's degree in computer science from Purdue University.

## Course Policies and Resources

### Readings and Recordings

There will be no required texts to purchase for this course but two references are recommended, both of which are available online for free through the library.

- J. Grus, *Data Science from Scratch: First Principles with Python*, 2019 (Second edition), O'Reilly (includes an introduction to Python, the language used for examples in the book and the course). \$29.99 on [Amazon](#) and free through the [CMU library](#)
- M. Cagan, *INSPIRED: How to Create Tech Products Customers Love*, 2018 (2<sup>nd</sup> edition), Wiley. \$21.38 for hardcover on [Amazon](#) and free through the [CMU library](#)

Online articles as listed in the course outline below will supplement these texts.

The Canvas site for this course is at <https://canvas.cmu.edu/courses/25163>. There is also a Piazza site linked to the Canvas site: Students registered for the class will be enrolled in Piazza as well; if you have not been enrolled, use the link [piazza.com/cmu/fall2021/95451](https://piazza.com/cmu/fall2021/95451)

Rather than emailing questions, students are encouraged to post on Piazza.

## Grading

There will be an initial Python Proficiency exercise worth 5% of the grade, followed by 3 homework assignments each due at midnight (Eastern Time) Each homework assignment will count for 20 percent of the grade, for a total of 60%. Late assignments (without a written excuse for medical/family/etc. emergencies) will be penalized at the rate of 10% of the assignment's grade per day late. A final project will count for 20 percent of the grade. The remaining 15 percent will be based on class attendance, participation in discussions, and occasional quizzes. One absence (with advance notice) is permitted, with further absences causing a proportional deduction in the class participation grade. There is no final exam for this class.

The code and results for homework assignments should be submitted in the form using Jupyter notebooks that we can run while grading. Use the file name convention `DSPM_HW<HW #>_<AndrewID>.ipynb`, e.g. `DSPM_HW2_steier.ipynb`. Within the Python code, file references should assume that the data file is in the same directory as the notebook (no hard-coded references to your personal directory structure), and keep the data file names as described in the assignment. You do not need to submit the data file. You may also submit a .pdf or word document writeup to accompany the notebook.

Grading will be on a straight scale as follows:

A+	98.0-100%	B+	88.0-89.9%	C+	78.0-79.9%
A	92.0-97.9%	B	82.0-87.9%	C	72.0-77.9%
A-	90.0-91.9%	B-	80.0-81.9%	C-	70.0-71.9%

Everyone taking the class should expect to register for a letter grade. Auditing the class or taking the class Pass/Fail are intended for extremely rare circumstances and only with consent of the instructor.

## Course Outline

This mini-course is planned around thirteen sessions of 1 hour 20 minutes apiece to be held Mondays and Wednesdays, 4:40 PM to 6 PM Eastern Time.

- **Week 1 (Aug 30 & Sep 1): Introductions; Metrics for Success**
  - **Topics**
    - Introductions
    - Data science and product management
    - Structure of the class
    - Metrics for success in product management
    - Customer acquisition and retention
    - Customer lifetime value

- **Readings:**
  - D. Yakobovitch, “The Intersection Between Data Science and Product Management,” November 11, 2020, <https://www.linkedin.com/pulse/intersection-between-data-science-product-management-yakobovitch/>
  - M. Cagan, *INSPIRED: How to Create Tech Products Customers Love*, “Part I Lessons from Top Tech Companies” <https://learning.oreilly.com/library/view/inspired-2nd-edition/9781119387503/p01.xhtml> and “Chapters 9-15: Product Teams”
  - Gorman, 2016, “A practical guide to calculating customer lifetime value,” <https://usermanual.wiki/Pdf/gormananalysiscomPractical20Guide20to20Calculating20Customer20Lifetime20Value20CLV.2014990060>
- **Assignment out**
  - Python proficiency exercise
  - HW1

## ● **Week 2 (Sep 8): Data for Product Management**

- **Topics**
  - Selecting data sources for product management
  - Exploratory data analysis
- **Readings/preparation**
  - Quartz, 2016, “Guide to bad data”, <https://github.com/Quartz/bad-data-guide>
  - Grus, J. Chapter 10, “Working With Data” (pp. 129-151) in *Data Science from Scratch*, 2019
- **Assignments due:** Python proficiency exercise (Sep 8)

## ● **Week 3 (Sep 13 & 15): Machine Learning for Customer Segmentation**

- **Topics**
  - Customer segmentation and machine learning
  - Supervised machine learning
  - Decision trees for segmentation
  - Unsupervised machine learning
  - Clustering for market segmentation
- **Readings:**
  - Grus. J., Chapter 11, “Machine Learning” (pp. 153-163), Chapter 17: “Decision Trees” (pp. 215-226) 2019 and Chapter 20: “Clustering,” (pp 263-277), *Data Science from Scratch*, 2019
  - L.A. Alzahrani, “Customer Segmentation: Unsupervised Machine Learning Algorithms In Python”, Jul 14, 2021,

<https://towardsdatascience.com/tagged/customer-segmentation?p=3ae4d6cfd41d>

○ Dancho, 2016, "Customer Segmentation, Part 1: K-means clustering," <http://www.business-science.io/business/2016/08/07/CustomerSegmentationPt1.html>

- **Assignment due:** HW1 (Sep 15)
- **Assignment out:** HW2

## ● **Week 4 (Sep 20 & 22): Analyzing Customer Feedback**

### ● **Topics**

- Introduction to natural language processing
- Social media listening
- Sentiment analysis

### ● **Readings:**

○ V. Yordanov, "Introduction to Natural Language Processing for Text," November 16, 2018, <https://towardsdatascience.com/introduction-to-natural-language-processing-for-text-df845750fb63>

○ Pandey, P., "Simplifying Sentiment Analysis using VADER in Python (on Social Media Text)" Sep. 23, 2018, <https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f>

○ Grus, *Data Science from Scratch*, 23, "Natural Language Processing," (pp. 279-293)

- **Assignment due:** HW2 (Sep 22)
- **Assignment out:** HW3, Final project

## ● **Week 5 (Sep 27 & 29): Web Analytics and Deep Learning**

### ● **Topics**

- A/B testing
- Clickstream analytics
- Deep learning

### ● **Readings**

○ VWO, "A/B Testing Guide," 2019, <https://vwo.com/ab-testing/>

○ Overgoor, "Experiments at AirBnB," 2014,

<https://medium.com/airbnb-engineering/experiments-at-airbnb-e2db3abf39e7>

○ Markou, 2017, "Clickstream analysis and data mining techniques 101: An introduction" <https://www.blendo.co/blog/clickstream-data-mining-techniques-introduction/>

○ 3Blue1Brown, "But what is a neural network?" Chapter 1 Deep learning," 2017 (video 20 minutes), <https://www.youtube.com/watch?v=aircAruvnKk>

○ (optional) Grus, J. Chapters 18-19 "Neural Networks", and "Deep Learning" (pp. 227-262) in *Data Science from Scratch*, 2019

- **Week 6 (Oct 4 & 6): Pricing; Managing Data Science Products**
  - **Topics**
    - Price elasticity and optimization
    - Managing AI / data science products
  - **Readings**
    - Meehan, Simonetto, Montan, and Goodin, Pricing and Profitability Management: A Practical Guide for Business Leaders, 2013. Chapter 1 available at <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-cons-pricing-and-profitability-management-chap1-excerpt-100314.pdf>
    - “What you need to know about product management for AI,” March 31, 2020, <https://www.oreilly.com/radar/what-you-need-to-know-about-product-management-for-ai/> . (The next two articles in the series are optional reading, but good too: <https://www.oreilly.com/radar/practical-skills-for-the-ai-product-manager/> and <https://www.oreilly.com/radar/bringing-an-ai-product-to-market/> ).
  - **Assignment due:** HW3 (Oct 6)
  
- **Week 7 (Oct 11 & 13): Final Presentations**
  - **Topics**
    - Final project presentations
  - **Assignment due:** In-class final project presentations (Oct 11 & 13) and reports (due Oct 13)

## Academic Integrity

Students are expected to strictly follow Carnegie Mellon University rules of academic integrity in this course. This means in particular that unless otherwise specified, homework are to be the work of the individual student using only permitted material and without any cooperation of other students or third parties. It also means that usage of work by others is only permitted in the form of quotations and any such quotation must be distinctively marked to enable identification of the student’s own work and own ideas. All external sources used must be properly cited, including author name(s), publication title, year of publication, and a complete reference needed for retrieval. The same work may not be submitted for credit in multiple courses. Violations will be penalized to the full extent mandated by the CMU policies. There will be no exceptions.

## Diversity

It is my intent that students from all diverse backgrounds and perspectives be well served by this course, that students’ learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, sexuality,

disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups.

## **Disability Accommodations**

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at [access@andrew.cmu.edu](mailto:access@andrew.cmu.edu).

## **Mental Health**

As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may diminish your academic performance and/or reduce your ability to participate in daily activities. CMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at <http://www.cmu.edu/counseling> . Support is always available (24/7) from Counseling and Psychological Services: 412-268-2922.