

# Spring 2021 Syllabus for 95-851: Making Products Count: Data Science for Product Managers

**NOTE:** The content below is from the Fall 2020 version of the course and will be updated for Spring 2021

**Spring 2021 Mini-course 1 (6 units):** MW 5:10 PM – 6:30 PM Eastern time.

**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:**

## 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
- 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 be given 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 co-instructs 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 two main references for this course on data science, and one on product management

- 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 <https://www.amazon.com/Data-Science-Scratch-Principles-Python/dp/1492041130>
- N. Joglekar and V. Nagaraj, *Digital Product Management Thinking: Integrating Analytics, Business Model, Coordination and Design Thinking*, SenseShaping Publishing, 2017. \$9.99 at [https://www.amazon.com/dp/B0741RQM7R/ref=rdr\\_kindle\\_ext\\_tmb](https://www.amazon.com/dp/B0741RQM7R/ref=rdr_kindle_ext_tmb)

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/19296> . 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/fall2020/95451](https://piazza.com/cmu/fall2020/95451) . Rather

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

All synchronous classes will be recorded via Zoom so that students in this course (and only students in this course) can watch or re-watch past class sessions. Please note that breakout rooms will not be recorded. I will make the recordings available on Canvas as soon as possible after each class session (usually within 3 hours of the class meeting). Recordings will live in our Canvas website. Please note that you are not allowed to share these recordings. This is to protect your FERPA rights and those of your fellow students.

## Grading

There will be an initial Python Proficiency exercise worth 5% of the grade, followed by 4 homework assignments each due at midnight (Eastern Time) Each homework assignment will count for 15 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) are permitted, with further absences causing a proportional deduction in the class participation grade. There is no final exam for this class.

Grading will be on a straight scale as follows:

A+	98-100%	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, is intended for extremely rare circumstances and only with consent of the instructor.

## Course Outline

This mini-course is planned around fifteen sessions of 1 hour 20 minutes apiece to be held Mondays and Wednesdays 5:10 PM to 6:30 PM Eastern time.

- **Week 1 (Feb 1 & 3): 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:**

- N. Joglekar and V. Nagaraj, “Chapter 1: Digital Project Management,” “Chapter 2: Business Context”, and “Chapter 9: Metrics,” in *Digital Product Management Thinking: Integrating Analytics, Business Model, Coordination and Design Thinking*, SenseShaping Publishing, 2017
  - Balfour, D. “Don’t Let the North Star Metric Deceive You,” January 2018, <https://brianbalfour.com/essays/north-star-metric-growth>
  - Gorman, 2016, “A practical guide to calculating customer lifetime value,” <https://gormananalysis.com/practical-guide-to-calculating-customer-lifetime-value-clv/>
  - (optional) Martin V., 2017, “Introduction to Product Management,” <https://www.udemy.com/product-management/> (2 hour free video course)
- **Assignment out**
  - Python proficiency exercise
  - HW1
- **Week 2 (Sep 9): Data for Product Management** (no class Sep 7)
  - **Topics**
    - Selecting data sources for product management
    - Exploratory data analysis
  - **Readings/preparation**
    - Grus, *Data Science from Scratch*, Chapter 10, “Working With Data” (pp. 129-151), 2019
    - Mawer, “The Value of Exploratory Data Analysis,” March 2017, <https://svds.com/value-exploratory-data-analysis/>
    - Quartz, 2016, “Guide to bad data”, <https://github.com/Quartz/bad-data-guide>
  - **Assignments due:** Python proficiency exercise (Sep 9)
- **Week 3 (Sep 14 & 16): Machine Learning and Customer Segmentation**
  - **Topics**
    - Customer segmentation
    - Decision trees for market segmentation
    - Clustering for market segmentation
  - **Readings:**
    - Cast, “Everything a Product Manager Needs to Know About Analytics,” 2013 <http://www.mindtheproduct.com/2013/02/everything-a-product-manager-needs-to-know-about-analytics/>
    - Grus. *Data Science from Scratch*, Chapter 11, “Machine Learning” (pp. 153-163), Chapter 17: “Decision Trees” (pp. 215-226), and Chapter 20: “Clustering,” (pp 263-277)
    - Dancho, 2016, “Customer Segmentation, Part 1: K-means clustering,” <http://www.business->

[science.io/business/2016/08/07/CustomersSegmentationPt1.html](https://science.io/business/2016/08/07/CustomersSegmentationPt1.html)

○ (Optional) Wind and Bell, 2008, “Market Segmentation”\_Chapter 11 in Baker and Hart, *The Marketing Book*, 2007 (sixth edition).

<https://marketing.wharton.upenn.edu/files/?whdmsaction=public:main.file&fileID=566>

- **Assignment due:** HW1 (Sep 16)
- **Assignment out:** HW2

- **Week 4 (Sep 21 & 23): Natural Language Understanding for Customer Feedback**

- **Topics**

- Introduction to natural language processing
- Text analytics to analyze customer feedback
- Social media listening

- **Readings**

- Grus, J. *Data Science from Scratch*, Chapter 23, “Natural Language Processing,” (pp. 279-293, remainder of chapter optional), 2019
- Ventislav 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)” Se[. 23, 2018, <https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f>

- **Assignment due:** HW2 (Sep 23)
- **Assignment out:** HW3

- **Week 5 (Sep 28 & 30): Evolving Products Based on User Feedback**

- **Topics**

- Product evolution
- Usability testing
- Survey and experimental design
- A/B testing
- Product instrumentation
- Clickstream analytics

- **Readings**

- N. Joglekar and V. Nagaraj, “Chapter 6: Product Definition and MVP Evolution,” “Chapter 7: Product Evolution,” and “Chapter 8: GTM,” 2017
- 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/>
    - Schuler, 2017 “10 product instrumentation mistakes and what we learned from them,” <https://sergioschuler.com/10-product-instrumentation-mistakes-and-what-we-learned-from-them-2f711d6030d3>
  - **Assignment out:** HW4, Final projects
- **Week 6 (Oct 5 & 7): Price Optimization and Deep Learning**
  - **Topics**
    - Price elasticity and optimization
    - Deep learning
  - **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>
    - 3Blue1Brown, “But what is a neural network?” Chapter 1 Deep learning,” 2017 (video 20 minutes), <https://www.youtube.com/watch?v=aircArvnKk>
    - . Grus, J. *Data Science from Scratch*, Chapters 18-19 “Neural Networks”, and “Deep Learning” (pp. 227-262), 2019
  - **Assignment due:** HW3 (Oct 7)
- **Week 7 (Oct 12 & 14): Data Science Scalability and Managing Data Science Products**
  - **Topics**
    - Parallel and distributed computing for scalability
    - Managing AI / data science products
  - **Readings:**
    - Wong, P and Bennett, R. , “Everything a Data Scientist Should Know About Data Management”, August 28, 2019, <https://towardsdatascience.com/everything-a-data-scientist-should-know-about-data-management-6877788c6a42>
    - “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 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:** HW4 (Oct 14)
- **Week 8 (Oct 19 & 21): Final Presentations**

- **Topics**
  - Final project presentations
- **Assignment due:** In-class Final project presentations (Oct 19 & 21) and reports (due Oct 21)

## Academic Integrity

Academic Integrity is expected at all time. Carnegie Mellon has an established well-defined policy on this subject which can be found at:

<http://www.cmu.edu/policies/documents/Academic%20Integrity.htm>

It is the responsibility of the student to verse themselves with these policies. All necessary and appropriate sanctions will be issued to all parties involved with plagiarizing any and all course work. Plagiarism and any other form of academic dishonesty that is in violation with these policies will not be tolerated. In particular, the same work may not be submitted for credit in multiple courses.

## Maintaining a Healthy Balance (as recommended and supported by the university)

Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress. All of us benefit from support during times of struggle.

You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful. If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support.

For local help and referrals, please contact the Office of the Dean of Student Affairs in 301 Warner Hall (412-268-2075). Counseling and Psychological Services (CaPS) at the Pittsburgh campus can also help you get connected to support. You can call them at 412-268-2922 and/or visit their website at <http://www.cmu.edu/counseling/> to learn more .

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

**CaPS: 412-268-2922**

**Suicide Prevention Hotline: 800-273-8255 (TALK)**