This is the syllabus for the 95-851 /451 course in Fall 23 Mini-2 and will be updated for Spring 2024

Fall 2023 Mini-2 Syllabus for
95-851: Making Products Count:
Data Science for Product Managers

Fall 2023 Mini-course 2 Section B2 (6 units) 3:30 PM -4:50 PM lectures on Zoom, except for Nov 13 and 15 which will be in-person in Hamburg 1007. Recitations on Fridays 11 AM – 12:20 PM in- person in Hamburg 1007
Professor: David Steier (steier@andrew.cmu.edu)
- Office hours: Immediately after class and on Fridays by appointment (send an email to schedule)

Teaching Assistants:
- Muhammad Asad Shoaib (mshoaib@andrew.cmu.edu)
- Ipsita Praharaj (iprahara@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. This course will introduce students 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 natural language processing.

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, and artificial intelligence as well as several executive education programs. 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. 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. David was also a Lecturer in Berkeley's Masters in Information and Data Science program, and was course lead for the Synthetic Capstone class. 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

Resources
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 the best companies create technology-powered products and services,* 2018 (2nd edition), Wiley. $26.72 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/37749 . There is also a Piazza site linked to the Canvas site: if you have not been enrolled, use the link https://piazza.com/cmu/fall2023/95851b2/home . 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 10 percent will be based on class attendance, participation in discussions, and 5% on a quiz given about halfway through the mini. Live attendance on Zoom is required for credit for class participation; limited exceptions will be made for medical reasons. 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.

Please include comments in your Jupyter notebooks to show what you've done. You need not do it for every cell, just where necessary. For those new to Jupyter, you could just use '#' to comment in the cells. You may use Markdown (go to Code drop down-->Markdown) to write your answers for reasoning-based questions (no need for hashtags here). You could even make use of HTML tags like <h1>,<h2>... for headers(size in decreasing order) and <br> for a new line.

Grading will be on a straight scale (no rounding up) 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 and will be taught on Zoom except for Nov 13 and 15, which will be taught in-person in Hamburg 1007. There are also recitations on Friday, which will be taught in person in Hamburg 1007, although a Zoom link will be provided. During the lecture sessions on Zoom, the classroom will be available for you to use during the class period, if needed. However, if you will be in the room, please make sure you use headphones to help control possible cross-feedback.

- **Week 1 (lectures Oct 23 & 25 on Zoom; recitation Oct 27 in HbH 1007): Introductions; Fundamentals of Product Management**
  - **Topics**
    - Introductions
    - Data science and product management
    - Structure of the class
    - Metrics for success in product management
    - Customer lifetime value
    - **Friday recitation**: Python review
  - **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/
    - "Part II: Product Teams" in M. Cagan, *Inspired* (2018)
    - https://shopify.engineering/a-data-scientist-s-guide-to-measuring-product-success
  - **Assignment out**
    - Python proficiency exercise
    - HW1
    - Final project

- **Week 2 (lecture Oct 30 and Nov 1 on Zoom, recitation Nov 3 in HbH 1007): Data for Product Management; Machine Learning for Customer Segmentation**
  - **Topics**

- Selecting data sources for product management
- Exploratory data analysis
- Customer segmentation and machine learning
- Supervised machine learning
- **Friday recitation**: Customer lifetime value
- **Readings/preparation**
  - Grus, J. Chapter 10, "Working With Data" (pp. 129-151
  - Grus, J. Chapter 11, "Machine Learning" (pp. 153-163), Chapter 17: "Decision Trees" (pp. 215-226) *Data Science from Scratch*, 2019
  -
- **Assignments due:** Python proficiency exercise (Nov 1)

- ## Week 3 (lecture Nov 6 & 8 on Zoom; recitation Nov 10 in HbH 1007): Machine Learning for Customer Segmentation; Demand Forecasting
  - **Topics**
    - Unsupervised learning for customer segmentation
    - Demand Forecasting
    - Time series analysis
    - **Friday recitation**: Machine learning for market segmentation
  - **Readings**:
    - Grus, J. 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 (Nov 8)
  - **Assignment out:** HW2

- ## Week 4 (lectures Nov 13 & 15 in Hamburg 1007; recitation Nov 17 in Hamburg 1007): Analyzing Customer Feedback
  - **Topics**
    - Introduction to natural language processing
    - Social media listening
    - Large language models
    - **Friday recitation**: Large language models for customer feedback
  - **Readings**:
    - V. Yordanov, "Introduction to Natural Language Processing for

Text," November 16, 2018, https://towardsdatascience.com/introduction-to-natural-language-processing-for-text-df845750fb63

- ○ T. Lee and S. Trott, "A jargon-free explanation of how AI large language models work," July 31, 2023, https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/
- ○ J. Grus, *Data Science from Scratch*, 23, "Natural Language Processing," (pp. 279-293)
- **Assignment out**: HW3

- ## Week 5 (lecture Nov 20 on Zoom, no recitation): Web Analytics and A/B Testing
  - **Topics**
    - ○ Quiz (Nov 20)
    - ○ Clickstream analytics
    - ○ A/B testing
  - **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/
    - ○ 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
  - **Assignment due:** HW2 (Nov 20)

- ## Week 6 (lectures Nov 27 & 29 on Zoom, no recitation): Pricing; Managing Data Science Products
  - **Topics**
    - ○ Price elasticity and optimization
    - ○ Managing AI / data science products
  - **Readings**
    - ○ A. Schur, "Demand forecast with different data science approaches", March 15, 2021, https://towardsdatascience.com/demand-forecast-with-different-data-science-approaches-ba3703a0afb6
    - ○ "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 (Nov 29)

- **Week 7 (Dec 4 & 6 on Zoom, no recitation): Final Presentations**
    - **Topics**
        - Final project presentations
    - **Assignment due**: Final project reports (Dec 8)

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.

HW#3 will encourage the use of generative artificial intelligence (AI) tools, such as ChatGPT. When AI use is permissible, it will be clearly stated in the assignment prompt posted in Canvas. For HW#1 and #2, use of generative AI is disallowed. For the final project, use of generative AI is permitted, but discouraged – it should be done without generative AI if possible. In assignments where generative AI tools are allowed, their use must be appropriately acknowledged and cited. For instance, if you generated the whole document through ChatGPT and edited it for accuracy, your submitted work would need to include a note such as "I generated this work through Chat GPT and edited the content for accuracy." Paraphrasing or quoting smaller samples of AI generated content must be appropriately acknowledged and cited, following the guidelines established by the APA Style Guide. It is each student's responsibility to assess the validity and applicability of any AI output that is submitted. You may not earn full credit if inaccurate on invalid information is found in your work. Deviations from the guidelines above will be considered violations of CMU's academic integrity policy. Note that expectations for "plagiarism, cheating, and acceptable assistance" on student work may vary across your courses and instructors. Please email me if you have questions regarding what is permissible and not for a particular course or assignment.

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.

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.