Masters in Data Science 2012-2019

The Master of Science in Data Science (MSDS) degree (formerly Master of Science in Analytics (MSAN)) launched its first cohort in the summer of 2012. From its inception, the MSDS program has strived to fulfill two distinct but complementary goals:

1. To collaborate with students on a rigorous exploration of the growing academic discipline of data science, and
2. To equip students with the practical tools and experience needed to secure employment in data science, particularly in the San Francisco Bay Area.

These goals are reflected in the program's vision and mission statements:

Program Vision: Our vision is to become a national leader in training the next generation of technically-competent and career-ready professionals who are fully engaged in, and continuously advance, the data science revolution in the Bay Area and beyond.

> ***Program Mission***: The mission of our program is to produce graduates who possess a theoretical and practical understanding of many classical and modern statistical modeling and machine learning techniques; who use contemporary programming languages and technologies to scrape, clean, organize, query, summarize, visualize, and model large volumes and varieties of data; and who use their knowledge and skills to successfully solve real-world data-driven business problems and to communicate those solutions effectively.

Program Learning Outcomes:

(1) Program Learning Outcome 1. Possess a theoretical understanding of classical statistical models (e.g., generalized linear models, linear time series models, etc.), as well as the ability to apply those models effectively.
(2) Program Learning Outcome 2. Possess a theoretical understanding of machine learning techniques (e.g., random forests, neutral networks, naive Bayes, k-means, etc.), as well as the ability to apply those techniques effectively of data.
(3) Program Learning Outcome 3. Effectively use modern programming languages (e.g., R, Python, SQL, etc.) and technologies (AWS, Hive, Spark, Hadoop, etc.) to scrape, clean, organize, query, summarize, visualize, and model large volumes and varieties of data.
(4) Program Learning Outcome 4. Be prepared for careers as data scientists by solving real-world data-driven business problems with other data scientists.
(5) Program Learning Outcome 5. Develop professional communication skills (e.g., presentations, interviews, email etiquette, etc.), and begin integrating with the Bay Area data science community.

As of this writing, the program is nearing the completion of cohort 7, and preparing for the arrival of cohort 8. This will be the program's first formal, outside academic review. In what follows we will outline the program structure and curriculum, faculty composition, student body and alumni, the practicum program, and the Data Institute; paying particular attention to challenges, both internal and external, facing the program.

Structure

The University of San Francisco (USF)'s MSDS Program is a twelve-month, full-time, academic program consisting of 35 units of instruction. Courses are offered at the University's downtown San Francisco campus (101 Howard Street) where it shares space with several other graduate programs including those offered by the School of Management (SOM), as well as University offices like Center for Institutional Planning and Effectiveness (CIPE), and the Alumni Engagement Office. (More on space challenges below.) Courses begin in July with an intensive bootcamp-style review, and continue throughout the year punctuated in seven weeks, half-semester modules. Courses are offered five days a week, during the day, and students are required to be available between the hours of 9am and 6pm (except during official University holidays and breaks). Starting in the second half of the fall semester and continuing through the remainder of the program, students are engaged in a curricular practicum course and spend two days per week working with a company on an industry data science problem, often at their respective company's location. More on the practicum program follows below.

In its original incarnation, the program was offered jointly by SOM and the College of Arts and Sciences (CAS), though now is strictly a CAS program. Notably, all undergraduate degree programs, and most CAS graduate programs, are located on USF's main campus located at 2130 Fulton St. This campus is often referred to as the Hilltop campus.

The current (and next) cohort represent the program operating at maximum capacity. Each course is run in two identical sections of around forty to forty-five students (depending on attrition). These sections are close to room capacity in the large fifth-floor classrooms which are the building's largest save the colloquium room on the first floor.

Curriculum

Students arrive at the program with a variety of academic backgrounds and work experience. Many students approach the program as a means of transitioning into a more technical career path, or as a pivot from a less technical educational background. In order to accommodate a wide variety of backgrounds the program perquisites are kept to a minimum and consist of:

1. A course in computation (Python or other high-level programming language).
2. A course in linear algebra.
3. A course in inferential statistics (need not be calculus based).

These prerequisites can be completed at any accredited university (online or otherwise). The current curriculum is laid out in Table 1, and next year's curriculum is found in Table 2. Previous iterations of the curriculum, as well as assessment materials like learning outcomes and curriculum maps are included in the Appendix. All students begin the program in the summer bootcamp where intense review courses are run in each of the three prerequisite areas. Students are assigned to two of the three courses based on their academic background and the results of the technical interview. (The technical interview is required of all incoming students as part of the admissions process. The interviews are conducted by program faculty and the candidates are evaluated in each of the three prerequisite subjects as well as communication potential.) In addition, every student takes an introductory course in basic data analysis with

relational databases using SQL. (The SQL class is new for cohort nine and was previously an EDA course focusing on an introduction to R.)

After the five-week bootcamp, the program begins in earnest in the fall. Each semester is divided into two modules lasting approximately seven weeks, during which students take either 6 or 7 units as designated in the tables below. Summer module electives this cohort include Deep Learning, Reinforcement Learning, and Natural Language Processing. Table 1 gives the current curriculum while Table 2 shows the updated curriculum taking effect in the 2019-2020 year. Course descriptions follow in the appendix.

Table 1

Course Name	Course Number	Units	Semester
DATA SCIENCE BOOTCAMP COURSES			
Exploratory Data Analysis	MSAN 593	1	Summer 2018
Plus complete two of the following:			
Computation for Analytics	MSAN 501	1	Summer 2018
Review of Probability and Statistics	MSAN 504	1	Summer 2018
Review of Linear Algebra	MSAN 502	1	Summer 2018
DATA SCIENCE SEMINAR COURSES			
Seminar Series I	MSAN 640	0	Fall 2018
Seminar Series II	MSAN 641	0	Fall 2018
Seminar Series III	MSDS 642	0	Spring 2019
Seminar Series IV	MSDS 643	0	Spring 2019
Seminar Series V	MSDS 644	0	Summer 2019
DATA SCIENCE REQUIRED COURSES			
Linear Regression Analysis	MSAN 601	2	Fall 2018
Time Series Analysis	MSAN 604	2	Fall 2018
Practicum I	MSAN 605	1	Fall 2018
Communications for Analytics	MSAN 610	1	Fall 2018
Intro to Machine Learning	MSAN 621	2	Fall 2018
Relational Databases	MSAN 691	1	Fall 2018
Data Acquisition	MSAN 692	2	Fall 2018
Distributed Computing	MSAN 694	1	Fall 2018
Distributed Data Systems	MSAN 697	2	Intersession 2019
Product Analytics	MSDS 603	2	Spring 2019
Data Visualization	MSDS 622	2	Spring 2019
Practicum II	MSDS 625	2	Spring 2019
Practicum III	MSDS 627	2	Spring 2019
Computational Statistics	MSDS 628	2	Spring 2019
Experiments in Data Science	MSDS 629	2	Spring 2019
Advanced Machine Learning	MSDS 630	2	Spring 2019
Problem Solving with Python	MSDS 689	1	Spring 2019
Special Topics in Analytics	MSDS 631	2	Summer 2019
Practicum IV	MSDS 632	1	Summer 2019

Table 2

Course Name	Course Number	Units	Semester
DATA SCIENCE BOOTCAMP COURSES			
Relational Databases	MSAN 691	1	Summer 2019
Plus complete two of the following:			
Computation for Analytics	MSAN 501	1	Summer 2019
Review of Probability and Statistics	MSAN 504	1	Summer 2019
Review of Linear Algebra	MSAN 502	1	Summer 2019
DATA SCIENCE SEMINAR COURSES			
Seminar Series I	MSAN 640	0	Fall 2019
Seminar Series II	MSAN 641	0	Fall 2019
Seminar Series III	MSDS 642	0	Spring 2020
Seminar Series IV	MSDS 643	0	Spring 2020
Seminar Series V	MSDS 644	0	Summer 2020
DATA SCIENCE REQUIRED COURSES			
EDA and Visualization	MSAN 593	1	Fall 2019
Linear Regression Analysis	MSAN 601	2	Fall 2019
Time Series Analysis	MSAN 604	2	Fall 2019
Practicum I	MSAN 605	1	Fall 2019
Communications for Analytics	MSAN 610	1	Fall 2019
Intro to Machine Learning	MSAN 621	2	Fall 2019
Data Acquisition	MSAN 692	2	Fall 2019
Distributed Computing	MSAN 694	1	Fall 2019
Machine Learning Laboratory	MSDS 699	1	Fall 2019
Distributed Data Systems	MSAN 697	2	Intersession 2020
Product Analytics	MSDS 603	2	Spring 2020
Practicum II	MSDS 625	2	Spring 2020
Case Studies in Data Science	MSDS 626	2	Spring 2020
Practicum III	MSDS 627	2	Spring 2020
Experiments in Data Science	MSDS 629	2	Spring 2020
Advanced Machine Learning	MSDS 630	2	Spring 2020
Ethics in Data Science	MSDS 633	1	Spring 2020
Data Structures and Algorithms	MSDS 689	1	Spring 2020
Special Topics in Analytics	MSDS 631	2	Summer 2020
Practicum IV	MSDS 632	1	Summer 2020

The curriculum has changed considerably with the evolution of the program (See supporting documentation). From its inception, the MSDS has positioned itself at the technical edge of

comparable programs in data science and analytics. At the same time, data science has grown dramatically as both an academic subject and a career path, and the Bay Area has been very much at the center of that development. The MSDS market positioning reflects our location in the Bay Area where technical skills are highly valued and highly sought after. The ongoing curricular changes in the program reflect the faculty's attempts to keep the program well calibrated to practical data science as it is understood and practiced in the Bay Area. In particular, this is reflected in the evolution of courses in the program including the addition of courses like Data Acquisition, Distributed Computing, Deep Learning, and Experiments in Data Science, and a reduction in the number of explicit soft-skills courses (with emphasis on those skills redistributed throughout the breadth of the curriculum). This is also reflected in the transition of the programs tech stack which continues to migrate towards Python, graphics processing unit (GPU), and cloud computing (primarily Amazon Web Services (AWS)), in order to better reflect the Bay Area tech industry.

Weekly Seminar

The program hosts a weekly seminar featuring industry speakers delivering lectures on topics of interest in applied data science. The seminar is advertised as a Meetup and is open to the public. Seminar runs each Friday afternoon at 12:30pm. The Meetup has over 8000 members and is well attended by local data scientists and students. MSDS students are required to attend as part of the MSDS curriculum. (The Seminar Courses listed in the program's curriculum above reflect this requirement.) Seminar also serves as a good opportunity for companies to advertise interest in hiring students, and students, attendees, and speakers often network after the lecture.

Student Body

The MSDS program has been successful in attracting a strong and diverse student body and has seen incredible growth over a short period of time. The original 2012-2013 cohort consisted of just 12 students.

The current cohort (2019) consists of 83 students from a starting point of 95 students in bootcamp. 55% of our student body is international, 53% of our student body identifies as female, and 24% of our current students have prior work experience. Countries represented include China, Colombia, Egypt, France, India, Iran, Japan, Malaysia, Mexico, New Zealand, Nigeria, Peru, Singapore, South Korea, Taiwan, and the USA. Average incoming GPA is 3.55 and the median GRE quantitative score is 166 (out of 170; approximately 90th percentile). The incoming 2019-2020 cohort will be similar in size and demographics to the current cohort.

Complete data is available from 2015-present and that information is summarized below.

Most Common Undergraduate Majors

- Mathematics (33%)
- Engineering (many disciplines) (31%)
- Economics (18%)
- Computer Science (11%)
- Statistics (10%)

Median Undergraduate GPA: 3.5

Median Quantitative GRE Score: 165

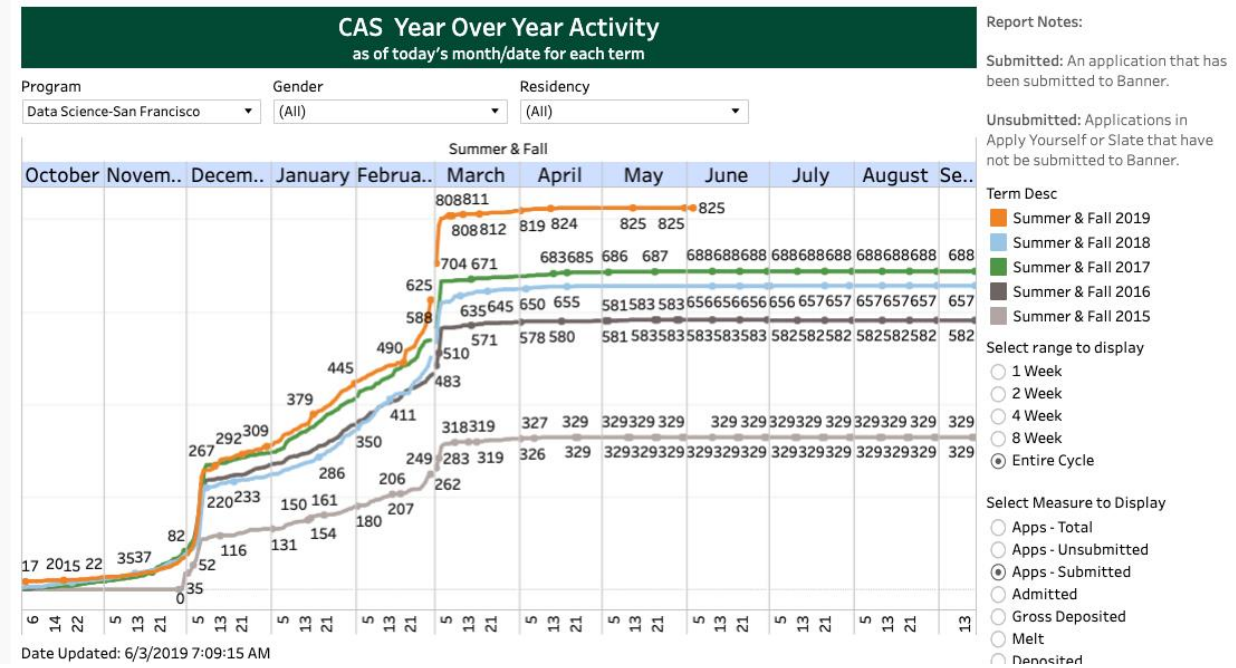
Median Quantitative GMAT Score: 50

Median Prior Work Experience (Years): 2

The MSDS program has had great success recruiting students from prestigious institutions, both nationally and internationally. Top schools that we consistently enroll students from since the program's inception in 2012 include UC schools like Berkeley, Davis, Irvine, Los Angeles, Santa Barbara, San Diego and Santa Cruz, and also New York University, Northwestern University, University of Illinois at Urbana-Champaign, University of North Carolina at Chapel Hill, University of Texas at Austin, University of Wisconsin, University of Washington, and the University of Southern California. Internationally, we enroll students from the Indian Institute of Technology, Peking University, Sun Yat-Sen University, University of Waterloo, Wuhan University and Zhejiang University.

Applications remain strong and the program saw more than 820 submitted applications for the 19-20 cohort. This represents a 26% increase from the previous year. See Table 3.

Table 3



As evidenced above, the students in the MSDS program are by and large outstanding and represent among the best students at USF across programs. Still, challenges remain around our ability to continue to recruit a diverse and talented class in the face of rapidly growing level of competition in data science education. In particular, the program's reputation and the option of STEM OPT extension upon graduation keeps international interest high.

There are, however, some early indication that domestic applicants are yielding at a lower rate than in years past. As noted in Table 3, the program has been able to increase our female enrollment by 80% since 2014 and female graduates earn salaries equal or greater to their male counterparts on average (see “Program Outcomes”). Still, we struggle to yield high quality female applicants, as over 60% of domestic female candidates who were offered admission in 2019 did not make an admission deposit. Additionally, despite an overall 26% increase in applications from 2018-2019, domestic applications remained flat. Domestic female deposits dropped by 27% and overall domestic deposits dropped by nearly 15% in this same time period. See Tables 4, 5 and 6.

One approach to recruiting strong domestic applications is forging 4+1 agreements with well-aligned institutions. These agreements guarantee admission to MSDS for students from partner institutions who have completed select courses and obtained the recommendation of their faculty. Current agreements exist with College of Charleston and St. Lawrence University (and of course, USF). Developing more such relationships is a priority moving forward.

Table 4

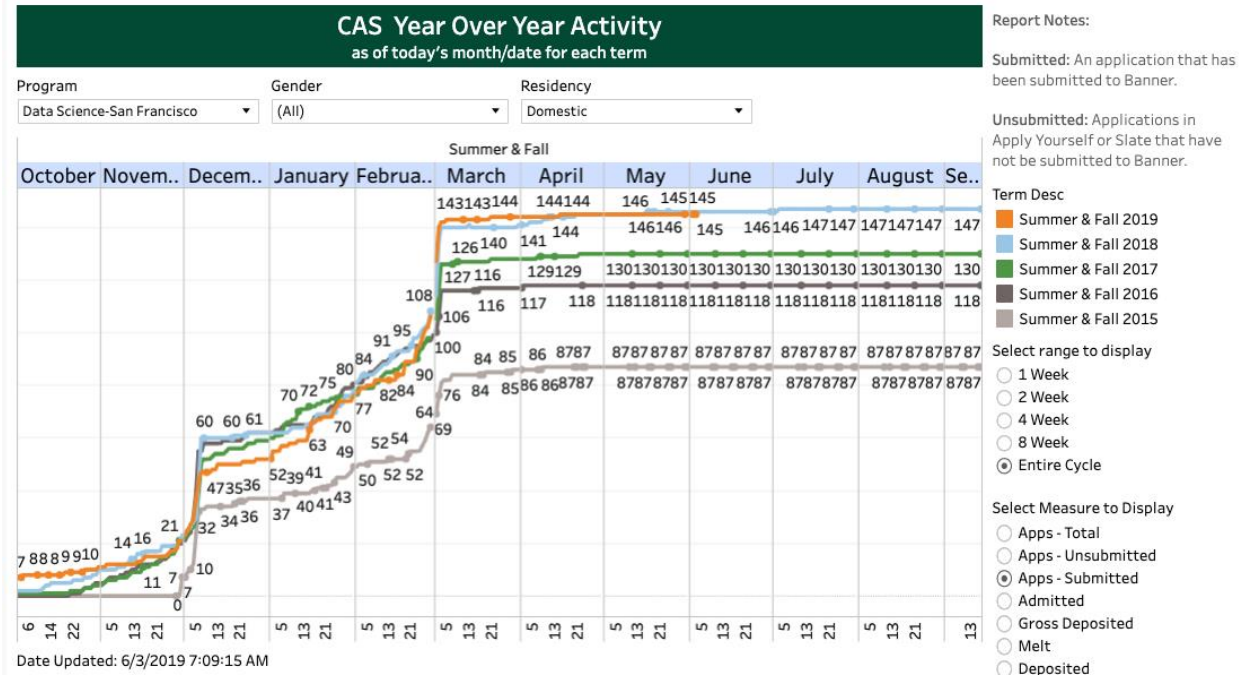


Table 5

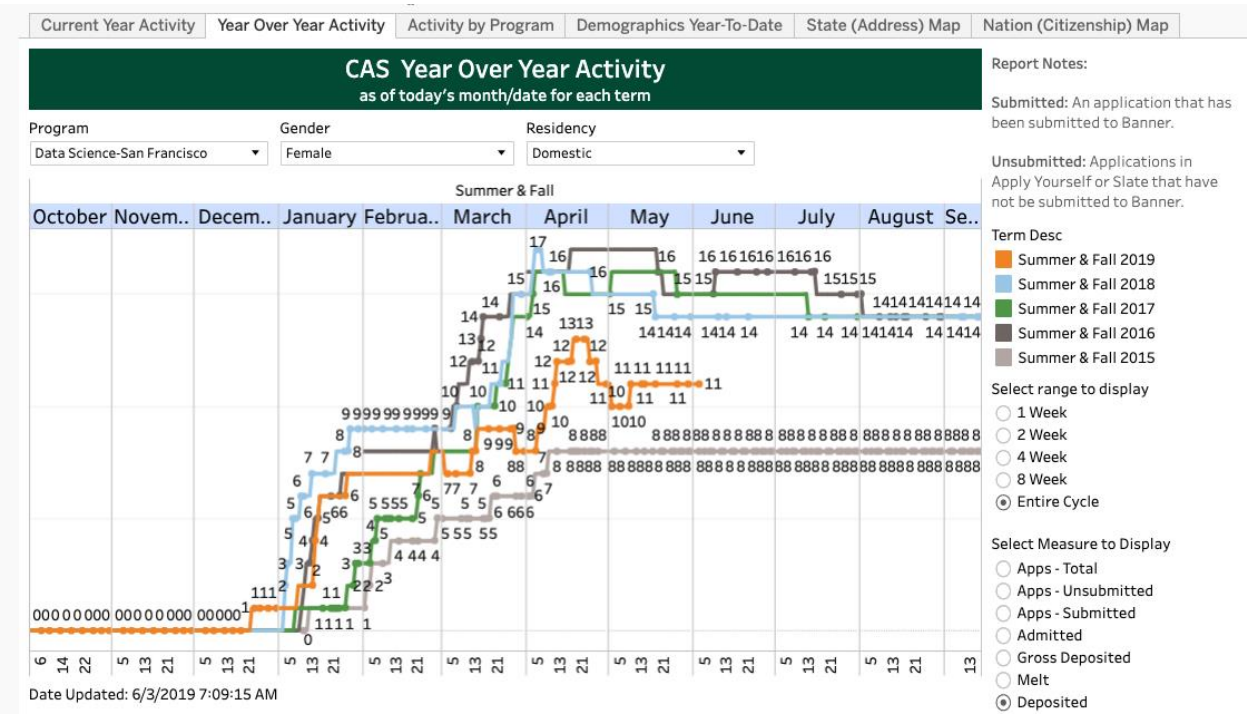


Table 6

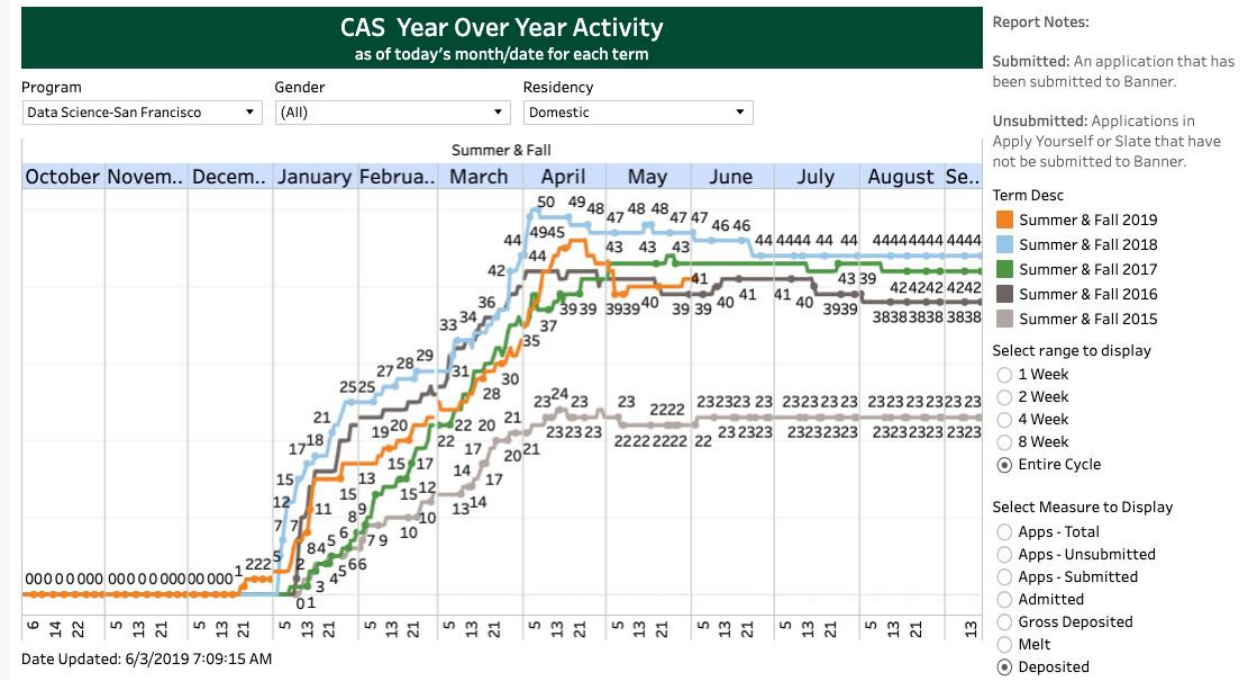


Table 7 below gives basic demographic information for each cohort thus far. More detailed information is in the appendix.

Table 7

Class of 2013	
Total	12
CITIZENSHIP	
US Citizen or Permanent Resident	9
Non-US Citizen	3
GENDER	
Male	7
Female	5

Class of 2014	
Total	23
CITIZENSHIP	
US Citizen or Permanent Resident	11
Non-US Citizen	12
GENDER	
Male	16
Female	7

Class of 2015	
Total	33
CITIZENSHIP	
US Citizen or Permanent Resident	22
Non-US Citizen	11
GENDER	
Male	22
Female	11

Class of 2016	
Total	41
CITIZENSHIP	
US Citizen or Permanent Resident	21
Non-US Citizen	20
GENDER	
Male	24
Female	17

Class of 2017	
Total	60
CITIZENSHIP	
US Citizen or Permanent Resident	34
Non-US Citizen	26
GENDER	
Male	32
Female	28

Class of 2018	
Total	80
CITIZENSHIP	
US Citizen or Permanent Resident	33
Non-US Citizen	47
GENDER	
Male	44
Female	36

Class of 2019	
Total	83
CITIZENSHIP	
US Citizen or Permanent Resident	37
Non-US Citizen	46
GENDER	
Male	38
Female	45

Faculty

The following faculty have some MSDS affiliation:

Tenured Full Professors: Terence Parr, Jeff Hamrick
Tenured Associate Professors: Stephen Devlin, David Uminsky
Tenure Track Assistant Professors: Nick Ross, Diane Woodbridge, Yannet Interian, James Wilson, Shan Wang, Paul Intrevado
Term Assistant Professor: Brian Spiering
Other: Rachel Thomas (position funded through the Data Institute)

Parr served as the inaugural director of the program. Hamrick took the mantle of director in January 2014 until his promotion to Vice Provost, at which point Uminsky became director in January 2015. Uminsky served as program director until January 2019 at which point Devlin took over as the current program director and Uminsky moved into his current role as director of the Data Institute (DI). More on the DI below.

The number and affiliation of faculty members requires clarification. In the early days of the program there was no MSDS department. As the program grew and generated faculty lines,

those lines were required to live in home departments. Moreover, in its original incarnation the MSDS program was shared between the College of Arts and Sciences (CAS) and the School of Management (SOM). SOM chose to end their affiliation with the program during its third year, and it moved completely into CAS. As a result, Intrevado moved from SOM to CAS, and his line was located in Department of Mathematics and Statistics (M&S). Devlin's and Uminsky's lines are entirely located in M&S, while Parr's line is in the Department of Computer Science (CS). Over time the home-department-requirement on lines was lifted, and Ross, Spiering, Woodbridge, and Interian were all hired directly into MSDS.

Wilson's and Wang's lines are also located in M&S, though half of their contractual teaching obligations lie in MSDS. The current agreement between M&S and MSDS is that these two faculty members alternate semesters between programs. For example, Wilson is spending the current spring semester (spring 19) at MSDS (teaching Computational Statistics (mod 1) and Experiments in Data Science (mod 2)). Wang will be at MSDS in the fall (teaching Linear Regression and Time Series Analysis) while Wilson teaches in the Bachelor's of Science in Data Science (BSDS) program on main campus. Wilson and Wang are the only two faculty with backgrounds in statistics. (Jeff Hamrick also has a statistics background, remains affiliated with the program, and even taught linear regression for the current cohort when a former faculty member left the University last year. However, his full-time responsibilities as Vice Provost mean that these engagements are, at best, one-offs.)

Thomas and Howard are primarily research scientists affiliated with the Data Institute and its certificate courses, though they do teach occasional courses in MSDS in what might be best described as an adjunct role. Thomas is teaching an elective course (Natural Language Processing) in the current cohort's summer module.

Other faculty have been affiliated with MSDS who are no longer part of the program. Nathaniel Stevens was a faculty member shared between MSDS and M&S. Stevens was a statistician (hired at the same time as Wilson) and an outstanding colleague who we were sad to lose when he decided to return as a faculty member to his alma mater, the University of Waterloo. Stevens' line is being replaced by Wang whose first semester at USF will be in the coming fall. David Guy Brizan was a member of MSDS who has since moved to the CS department (though he continues to occasionally mentor practicum projects for MSDS). Brizan's line was replaced by Spiering (in a term teaching role) following a failed search for a tenure track faculty member.

The discussion above brings us to a considerable challenge facing the program. First, staffing is thin and redundancy is nearly nonexistent. If Wilson or Wang, for example, take a sabbatical leave, the program has no capacity to absorb their loss and continue to deliver the required courses in both MSDS and BSDS. When Terence Parr recently went on sabbatical, faculty members in the program covered the loss via overload teaching. One faculty member did in excess of 28 units of teaching to accommodate Parr's sabbatical—this equates to 10 units of overload teaching, or the equivalent of 2.5 extra courses. While admirable, such overload teaching inevitably interferes with a faculty member's other contractual obligations including research and service. Uminsky delayed his own post-tenure sabbatical to continue serving in the role of program director.

Secondly MSDS draws significantly from faculty in other departments to fill its needs. This places great strain on the affected home departments, with the M&S department and BSDS programs feeling that strain most acutely. Examples of this borrowing include the loss of a term faculty position in M&S that had initially offset the loss of David Uminsky to the MSDS program. With the appointment of Steve Devlin as Uminsky's replacement as program director, M&S is

currently operating in full crisis mode minus the usual teaching contributions of Devlin and Uminsky (and for the foreseeable future). For example, the M&S department recently hired Daniel O'Connor to teach in the BSDS program and to fill the line vacated when Xuemei Chen left the University to solve a two-body problem. In the spring semester, Devlin was asked to take over as the MSDS program director. Devlin, however, was slated to teach Introduction to Statistical Learning (Math 373) on main campus in the BSDS program. When he moved downtown, O'Connor agreed to take the class over. O'Connor's willingness to do so, and flexibility, are admirable, but left him with two brand new upper division course preps in his first year. (He also taught Math 372, Applied Statistics.) Normally, a first-year new hire would get a more forgiving teaching schedule.

Compounding these issues, MSDS has had very little success finding competent adjunct instructors with the skills to teach the technical and specialized courses in the curriculum during the scheduled times. Timing is especially rigid because of the need to accommodate practicum twice per week, and the fact that courses run during the day. In addition, the CS department has had only minimal engagement with MSDS outside of Parr. This is atypical of a data science program to say the least, and is due in part to the fact that undergraduate CS enrollments have grown dramatically, and CS runs its own graduate program on main campus. Still, it's a considerable frustration when MSDS is, for example, unable to staff its data visualization course despite the presence of several visualization experts in the CS department. This was precisely the situation in the spring 2019 semester.

Faculty are certainly entitled to a preference regarding their teaching and program engagements. However, the commitment by the University to run a high-quality graduate program inevitably requires logistical considerations and careful utilization of scarce faculty resources. It is not clear that University administration is willing to get involved in discussions surrounding difficult questions of how faculty resources should be organized around MSDS and closely related departments. It's also worth noting that the union structure in place at USF means that chairs and program directors have no power to make teaching assignments or compel faculty to non-teaching commitments. Adding to the complication is the difficulty of the geographical separation of MSDS from main campus. The downtown campus is only located 3.7 miles from the main campus. However, travel time is often in excess of 45 minutes with traffic. To be clear, the downtown location of the MSDS program is essential to its identity and ability to recruit students. At the same time, however, the challenge of commuting between campuses leads to a situation where many faculty prefer one campus over the other. Without an administrative liaison, unsustainable pressure builds on program directors and chairs. Moreover, faculty who are commuting between campuses and teaching in both programs feel stretched in a way that colleagues who simply opt out are not. Finally, the fact that junior MSDS faculty are located at the downtown campus results in an unfortunate level of separation from the rest of the University. These faculty have far less interaction with their colleagues and the deans, who work almost exclusively on main campus, and as a result have less presence on University committees. These are concerning issues for junior faculty who will be peer-evaluated for tenure and promotion.

BIOS

***Stephen Devlin**:* Stephen is an Associate Professor in the Department of Mathematics and Statistics and is the director of the Master's in Data Science program (MSDS). He has also served as department chair and director of the undergraduate data science program. Stephen has a bachelor's degree from Manhattan College in New York, and a Ph.D. in mathematics from the University of Maryland. He was a C.L.E. Moore Instructor at MIT before moving to the University of San Francisco. His research interests include both pure and applied mathematics and he has published in diverse fields including representation theory, game theory, and sports analytics.

***Jeff Hamrick**:* Jeff has been instrumental in developing USF's MS in Data Science (MSDS) program. His teaching also supports the university's Master of Science in Financial Analysis (MSFA) program. For the MSDS program, he has designed courses in linear regression analysis, time series analysis, multivariate statistical analysis, and SAS programming. In the MSFA program, he teaches econometrics and financial econometrics. Professor Hamrick contributes his extensive knowledge of finance, mathematics, and statistics to prepare students for both traditional jobs in finance and cutting-edge jobs related to data science. Professor Hamrick has experience in hedge fund management and consulting for both financial services and software engineering firms. He is a CFA charterholder and a chartered Financial Risk Manager (FRM), with publications in nonparametric statistics and mathematical finance. Prof. Hamrick's academic work in "big data" stems from his interests in natural language processing, market microstructure (e.g., models of intraday asset prices), and sabermetrics (i.e., the use of statistical and computational methodologies to study baseball).

***Yannet Interian**:* Yannet is an Assistant Professor at University of San Francisco. Her research interests are data mining and applied statistics, using large datasets. She solves problems from advertising, viral media and on-line social networks. She holds a PhD in Applied Mathematics from Cornell University and a BS in Mathematics from the University of Havana, Cuba. She was a postdoctoral fellow at UC Berkeley. She worked for five years as a Data Scientist at Google. She co-founded Akualab, a start-up that helped organizations develop data-driven products using machine intelligence. She has designed data science courses for Berkeley and USF and is currently teaching full time for USF's Master's in Data Science program.

***Paul Intrevado**:* Paul Intrevado is an Assistant Professor of Data Science in the Department of Mathematics & Statistics. Paul is a co-director of the Machine Learning, Artificial Intelligence, Gaming Intelligence and Computing at Scale (MAGICS) Lab. He previously served as the MS Data Science Practicum Director [2014-2017], and is the founding Associate Director of the Data Institute [2017].Professor Intrevado's research focuses on service operations, including healthcare, hospitality and sports. Most recently, Prof. Intrevado's research has focused on the mathematical modeling and analysis of patient-centric long-term care network design, revenue management of food

and beverage services at large hotels and convention centers, as well as joint sabermetric analysis with the Houston Astros. Paul brings to USF a diverse wealth of professional experience from Canada, the USA, Kenya, Turkey, Norway, and India. He has consulted for the Veteran's Affairs Center for Applied Systems Engineering in Washington, D.C. and Palo Alto, for the AMPATH program in Eldoret, Kenya, to reduce and treat the spread of HIV/AIDS, as well as being a Research Fellow at the Montréal Jewish General Hospital in Montréal, Canada. Prof. Intrevado has won multiple teaching awards from both Santa Clara University's Leavey School of Business and McGill University's Desautels Faculty of Management. He is bilingual in both English and French.

***Terence Parr**:* Terence is a professor of computer science and data science at the University of San Francisco where he continues to work on his ANTLR parser generator. Until January 2014, Terence was the graduate program director for the computer science and was founding director of the MS in Analytics (now MS Data Science). Before entering academia in 2003, he worked in industry and co-founded jGuru.com. Terenceherded programmers and implemented the large jGuru developers' website. Terence has consulted for and held various technical positions at companies such as Google, Salesforce, Sun Microsystems, IBM, Lockheed Missiles and Space, NeXT, and Renault Automation. Terence has been an expert witness for Google twice, including for the Oracle v Google (Android/Java) trial. Terence holds a Ph.D. in Computer Engineering from Purdue University.

***Nicholas Ross**:* Nicolas received his PhD from UCLA's Anderson School of Business in 2012. His research interests revolve around how asymmetric information affects decision-making and, professionally, he builds systems that attempt to violate the CAP theorem. Currently, Nick works in the video game industry where he takes the fun out of games. Dr. Ross has many years of analytical experience all along the accounting spectrum. He is interested in focusing on how asymmetric information effects decision-making. The academic and job experience he has brings a lot to the USF table, especially because most of his industry work has been located right here in San Francisco. Dr. Ross worked as a litigation consultant at Bates White in Washington DC and San Diego, where he answered all math related requests for lawyers. Prior to joining USF, he spent over four years working in video games at two San Francisco based companies. At TinyCo, an Andreessen-Horowitz backed startup, he led the Analytics and User Acquisition teams and at Sega, he was Director of Analytics. Other than teaching, Dr. Ross is working on a piece called, *Does an Executive's Ability to Profit from Insider Trading Translate into Superior Form Performance?* The time Dr. Ross has spent on his research, education and job proves that he is making a difference and changing our world. He has developed curriculum for the Master of Data Science program and provided instruction in cost accounting at the undergrad level.

***Brian Spiering**:* Brian Spiering is an Assistant Professor at the University of San Francisco. His interests include software engineering, Natural Language Processing (NLP), Machine Learning (ML), and Artificial Intelligence (AI). He is the Career Services Advisor for their Masters in Data Science program. He earned a Ph.D. in Psychology

from the University of California, Santa Barbara. Brian is active in the San Francisco tech community through volunteering and mentoring.

***Rachel Thomas**:* Rachel Thomas has a math PhD from Duke and was selected by Forbes as one of "20 Incredible Women Advancing AI Research."She is co-founder of fast.ai and a researcher-in-residence at the University of San Francisco Data Institute, where she teaches in the Master of Science in Data Science program. Her background includes working as a quant in energy trading, a data scientist + backend engineer at Uber, and a full-stack software instructor at Hackbright. Rachel's writing has made the front page of Hacker News six times, has been translated into Chinese, Spanish, & Portuguese, and has been featured in newsletters from O'Reilly, Fortune, Mattermark, Metis, & others. She was a keynote speaker at JupyterCon, a featured speaker at PyBay, and writes an ask-a-data-scientist column at fast.ai.

***David Uminsky**:* David Uminsky is Associate Professor of Mathematics and current Executive Director of the Data Institute. He was the program director of the MSDS program from Fall 2014-Spring 2019. His research interests are in applied mathematics. He is interested in unsupervised machine learning, data clustering, algebraic signal processing, as well as pattern formation, dynamical systems and fluids. David was selected in 2015 by the National Academy of Sciences (NAS) as a Kavli Frontiers of Science Fellow. Each year, 100 researchers under the age of 45 are selected by the academy, and the 20% of the current NAS were previous Kavli Fellows. He is also the founding director of the Bachelor's program in Data Science at University of San Francisco. Before joining USF, he was a combined National Science Foundation and UC President's Fellow at UCLA, where he was awarded the Chancellor's Award for post-doctoral research. This award is given to approximately top 20 post-docs out of over a thousand who qualify for consideration. He holds a PhD in Mathematics from Boston University and a BS in Mathematics from Harvey Mudd College.

***James Wilson**:* James is an Assistant Professor of Statistics and Co-Director of the BS in Data Science program at the University of San Francisco. He has joint appointments in the Department of Mathematics and Statistics and the MS in Data Science program, where he has developed and taught courses in Bayesian statistics, machine learning, data science, and network analysis. In research, James develops new statistical and computational techniques to model, analyze, and explore high-dimensional and relational (network) data. Driven by applications in neuroscience and political and social science, James seeks to understand and motivate the interplay between statistics, data science, and application. James also works closely with companies in the Bay area to solve exciting data science and network analysis problems. He has worked with companies including Airbnb, Eventbrite, the San Francisco 49'ers, the Houston Astros, Xoom, and Zipcar. At the University of San Francisco, James seeks to promote the next generation of cross-disciplinary research among data scientists in academia, industry, and government. As a first step to accomplish this aim, he recently organized and chaired the first and second Data Institute Conferences, which brought together nearly

200 renowned data scientists from all three sectors, and plans to continue holding this conference annually.

Diane Woodbridge
Diane Woodbridge is an assistant professor in the MS in Data Science program at the University of San Francisco. Her research interests include database management systems, data fusion and data mining in various domains including biomedical, geoscience and geospatial remote sensing. Prior to joining USF, Professor Woodbridge was with the scalable analysis and visualization department at Sandia National Laboratories.

Staff

MSDS is lucky to have outstanding staff. Kirsten Keihl serves as the Assistant Director of the MSDS program; Mindi Mysliwiec is the Director of Operations of the Data Institute; and Leslie Blakeman is the MSDS Program Assistant IV. All three staff members are shared with the Data Institute with 2/3 of their duties in MSDS and 1/3 in DI.

One of the most labor-intensive aspects of the program is the admissions process. The early application deadline is mid-December and the final deadline is early March. The admissions process runs from September through May when final class composition is determined. Staff are responsible for, and admirably perform the related duties of preliminary application review and categorization. Applications that meet the preliminary consideration requirements are advanced to the interview stage. Twenty minute technical interviews are scheduled by staff in recurring time blocks offered by faculty. This year, more than 830 applications were reviewed and 303 interviews were conducted. This represents an incredible expenditure of time and effort.

Kirsten, Mindi, and Leslie are also essential in the day-to-day student-facing operation of the program. They are often the first line of communication when students have issues or concerns. In addition, they handle the logistics and organization of information sessions, marketing campaigns, communication with University offices, scheduling of internal and external events, program and certificate course informational requests, certificate course logistics (through the Data Institute), handling of practicum non-disclosure agreement (NDA), handling of student Curricular Practical Training (CPT) / Optional Practical Training (OPT) letters, etc.

Practicum

A unique feature of MSDS is the practicum: a part-time internship that every student is required to participate in. The practicum takes the place of a more traditional master's thesis, and is better aligned with the particular goals of MSDS, especially in light of the short duration, and industry focus of the program. In broad strokes, the students spend Monday and Wednesday working with a company on a data project. Practicum begins in the fall, and lasts for nine of the

twelve months of the program. Students have some choice in which project they work on, with a few caveats, and during the course of the practicum meet weekly with a faculty mentor who oversees the project and ensures that it achieves the appropriate curricular goals.

There are two critical success factors for the practicum program. The first is that the students work on a real data science project. Data science is a broad area and practicum projects reflect this breadth. While some projects may be incredibly technical, like deep-learning medical imaging projects, others may be focused more on data visualization or traditional customer analytics. Regardless, it is essential that students come away with industry experience working on a data-driven problem. This work becomes an important component of the student's resume, a talking point as they interview for jobs, and allows them to differentiate themselves from other applicants at a similar level of education and experience.

The second critical success factor of the practicum is that it be low-burden on the host organization. The MSDS faculty and administrators do everything possible to make the company's experience with the practicum students as frictionless as possible. The practicum director (currently MSDS faculty member Nick Ross) is incredibly responsive and accessible, making time for meetings year-round and providing a single point of contact if any confusion or issues arise. Weekly faculty mentoring meetings ensure that students have a way to ask questions about tools and technologies that they may not be familiar with. The faculty and staff have also created templates and examples for any required forms or documentation.

Implementation Specifics

Getting information to students

Around the mid-point of the fall module (roughly the middle of September), the students receive a pitchbook containing short project descriptions from each participating organization. These, roughly one-page pitches contain basic information about the organization and the people that the students will work with. Information about the project itself is kept deliberately high-level as projects tend to change focus over the course of the year. Organizations are specifically told to consider the project descriptions as more of a taste of the type of work a student will do rather than a complete description or contract.

In the following weeks, MSDS hosts a series of pitch-events, where companies give five-minute talks on their proposed project, and then stay for an hour of networking. Students are required to attend all of the pitch events, which in the 2018-2019 cohort was 16 hours in length spread over three days:

- Monday, September 24th 1-8PM
- Tuesday, September 25th 10:15 – 5PM
- Friday, September 28th 9-11AM

Each day is broken into two-hour segments, the first hour containing 7-10 organization pitches and the second hour reserved for networking time. Students are required to take these events seriously and present themselves professionally. During the 2018-2019 event roughly 60 companies and organizations pitched to the students. In 2017-2018 over 70 companies attended. The final pitch day is reserved for remote projects not located nearby. In this case, organizations present via videoconference and take questions from the student body which is gathered together in a single seminar room. Once the students have seen all the pitches, they fill out a survey where they choose their top-15 and bottom-5 projects.

Matching Students to Organizations

Of course, project demand is uneven among students and a matching has to occur. The practicum director, working closely with the faculty, makes preliminary matches as follows.

There are three primary considerations: (1) the organizations needs and requests, (2) student's preferences, and (3) students performance in classes to date. Organizations tend to ask for students with certain characteristics, such as experience in a particular industry or a specific skill set (like experience with a particular technology).

Matching is then done in stages, with 3-4 students proposed to each company with the choice: pick two, pick a subset to interview, or ask for different students (and state why). Every organization is different, but most choose to interview a number of students before deciding on a pair. For the students, this part of the process is incredibly stressful, and many of them will interview at their preferred organization only to be rejected. Still, over 75% of the students have been placed in one of their top 10 choices, and very few are placed at an organization not among their top 15.

The goal of the matching is to put students in a position to succeed. Once all students are matched with companies, faculty mentors are assigned to each project based on the faculty's area of expertise. For example, if a project requires NLP, faculty with NLP experience will be the mentor. Students are generally placed at organizations in pairs, though some are sent as singles and occasionally a triple is placed.

There are two process pieces of the matching that add difficulty. First, companies often require the use of an NDA or intellectual property (IP) agreement between a subset of USF, the company, and the student. This requires coordinating with USF's legal team and pushing them to complete tasks efficiently so that students can begin working in a timely manner. Secondly, international students require a CPT letter to be signed by an organizational representative and placed on letterhead. International students cannot begin working at their organization until their CPT letter has been processed, which can take up to a week. If a student violates their CPT arrangements they can lose their visa status, requiring them to leave the country. The MSDS staff organize both of these sets of documents for the students, working with the practicum director to make sure that USF is compliant with both the organization and the Immigration and Naturalization Services.

Students can begin work as early as the second week of October (when fall module two begins), but interviewing and placement frequently continue through the module. The most common project start date is early November, but final matches sometime stretch into the new year.

For the remainder of the program, students work two days a week on the project, meet with their mentors once a week, and fill out a weekly practicum update form which is sent to the faculty mentor, practicum director, and the company.

Faculty Roles and Responsibilities

Faculty work for the practicum begins in February of the preceding year, at which time the practicum director meets with the MSDS program director and other parties to put together a plan for the upcoming year. This plan includes identifying potential partner organizations,

deciding which relationships should be expanded, which should be closed down, and identifying any other issues that arose in the past year or can be anticipated in the upcoming year.

At this time, an initial list of two to three-hundred organizations is created. Each organization is vetted based on contacts at the organization, possible alumni presence, and estimate of student interest in working on a project at that organization. Between February and pitch day, the practicum director will email, call, and meet in-person with many of the organizations. Depending on the organization, we may invite them to MSDS events or provide them with tickets to our Data Institute Conference in order to cultivate the relationship with USF. The list of potential practicum companies also expands during this time as information about new potential partners becomes available.

The most work intensive time for the practicum director are the weeks leading up to pitch day, through the last student placement. During this time, an incredible variety of things occur, few of which are under the control of the practicum director. A few examples: In 2017, two companies essentially went out of business between pitch day and project launch. In 2018, two companies, both highly desired by students, backed out between pitch day and project launch. One of these was due to fit—they decided that USF students didn't have a sufficient background in their domain area—the other was due to a higher-up within the company initiating a last-minute veto. Since USF doesn't require organizations to sign an official contract to participate, it is impossible to get a full read on which organizations are going to participate, making asset allocation difficult. For example, one insurance company met with the practicum director on multiple occasions for over three hours total, only to decide against engagement.

Within USF's tracking documents, organizations are labeled on a five-point scale from "No" to "Likely." Faculty mentors on each project undertake two primary responsibilities for each of their practicum teams. The first role is that of mentor: making sure that the students are succeeding at the practicum and that the project is proceeding in reasonable manner. In this role, there are two common issues: weak projects where the students are doing something that isn't data science related. For example, a few years ago, a project ended up having a significant data entry component. At this exchange, the practicum director met with the company to reset expectations make sure that the project grew to include more sophisticated goals. The second part of this role is to provide guidance to the students regarding expectations in professional settings. Students in the MSDS program come from a variety of backgrounds and experiences and are often unaware of typical expectations around working in a professional setting. The faculty mentor can provide another perspective on each situation within a practicum, to help gauge if a specific situation is out of the ordinary or not. For example, a common situation is that the company mentor may not be available to meet for a few weeks at a time. One particular example of this occurred when the company mentor's fiscal plans were due at the end of March, and became the person's all-consuming focus. The students raised concern to the faculty mentor who was able to provide perspective, make sure that the students knew that this was a common occurrence, and advised the students to simply hang on for a week or two until the fiscal plan was complete. In each of these settings, the faculty mentor is not required to solve the problem, but rather highlight that an issue is occurring, and pass it on to the practicum director.

Once a project is well defined and set in motion the needs of the project itself subside, and the faculty mentor has a secondary role: career advisor. In this capacity, which usually starts around mid-April, the faculty actively work with the students on their job plan, interpreting interviews, resume reviews and negotiating salaries. The hope is that the faculty mentor will have a strong bond with the students, so that the students will listen to their advice. Similar to the previous

role, the faculty mentor provides advice and a guidance, and alerts the program director if anything is out of the ordinary with a student.

Staff are also incredibly important to the smooth functioning of the practicum. As previously mentioned, they handle the legwork of 85 NDAs and around 50 CPT letters, making sure that all are filed properly. The staff also work to set up the pitch day events, reserving rooms, scheduling catering, and handling a huge variety of logistical issues. Finally, the staff also does an incredible amount of support work, such as sending emails to schedule the 60 organizations coming in to pitch, editing the pitches that each organization sends, taking pictures, and sending thank you notes.

Program Employment Outcomes

Program outcomes are strong. Given the industry focus of the program, the primary metric for evaluation is employment and almost all students receive offers before or shortly after completion of the program. See Table 8 below.

Table 8

Graduation Year	Offers at 6 Months	Median Base Salary	Max Salary	Most Common Job Title	Employed by Practicum Company
2018	100%	$120,000	$177,000	Data Scientist (64%)	15%
2017	98%	$110,000	$180,000	Data Scientist (49%)	13%
2016	100%	$110,000	$125,000	Data Scientist (56%)	15%
2015	100%	$100,000	$160,000	Data Scientist (47%)	26%
2014	100%	$90,000	$160,000	Data Analyst (43%)	4%
2013	100%	$85,000	$100,000	Data Scientist (33%)	0%

Median starting salaries remain consistent across demographics, with women and international students earning salaries equivalent to those of all graduates. Female and international graduates in 2018 earned a median base salary of $120,000. Graduates with no prior work experience earned a median base salary of $115,000. The return on investment of this program remains very high as graduates who were working before entering the program increased their median base salary by an average of $50,000 after graduation.

Since 2012, 52% of graduates were hired with the title of data scientist, 14% as data analysts, and 8% as machine learning or data engineers. The top companies that hire from our alumni pool include Amazon (9), Apple (5), Facebook (9), Google (6), Pinterest (4), Trulia (4), Uber (5), and Walmart (4). Graduates continue to advance in their careers as 17% of graduates now hold the title of senior, manager, director or lead.

Alumni

As detailed above, graduates from the MSDS program begin or continue careers as data scientists, analysts and engineers and enjoy one of the highest employment rates of any graduate data science program in the country with 93% of graduates of all seven MSDS cohorts receiving a full time offer of employment within three months of graduation. In addition, 97% of

all graduates of our most recent 2018 cohort received a full time offer of employment within three months of graduation.

Alumni remain engaged after graduating through events hosted by the program including bi-annual alumni happy hours, five-year class reunions, participating in the practicum program as company hosts and mentors, hosting bi-weekly alumni lunches and mentoring sessions, speaking at our Seminar Series, teaching and grading for classes, speaking on panels for MSDS and DI events, and serving as board members on the DI Advisory Board. Alumni also serve as a valuable resource to current students and fellow alums as many use our program as a hiring pool and frequently share employment opportunities and offer referrals. Once per year, once students have begun their job search, around 20 alumni participate in mock interviews with our current students and return to campus to interview students, offer advice and feedback. Alums also give back financially, making gifts to the MSDS scholarship fund through the Data Institute each year for the “Day of the Dons” annual giving campaign.

Overall program satisfaction levels remain high, as reported by graduating students in our annual exit survey. In 2018, 98.7% of graduates were satisfied or very satisfied with their experience in the MSDS program and 96% would recommend the program to a qualified friend. Graduates were satisfied with the quality of support offered by MSDS faculty as 97% of students reported being satisfied or very satisfied (47% very satisfied), and 93% were satisfied or very satisfied with the quality of teaching in the program. 97% were satisfied or very satisfied with the career services offered by the program, and 94.74% found their faculty mentor to be a helpful resource for solving technical problems for practicum but also for professional and career advice.

Admissions

The admissions process for MSDS is particularly labor intensive. Applications have grown each year with over 820 applications for the 2019-2020 cohort. As noted above, applicants are required to have completed (or be enrolled in) prerequisite classes in statistics, linear algebra, and programming from an accredited college or university. In addition, applicants are required to take the GRE general or the GMAT test, and TOEFL where applicable. Applications are first screened by MSDS staff who review for completeness, review test scores, check transcripts, and read statements of purpose. Applicants who pass this initial screen are placed in an interview queue. There, the applicant chooses from available time slots for a 15 to 20-minute video conference interview with an MSDS faculty member. Each MSDS faculty member sets aside weekly blocks for these interviews during the interview season. (The early deadline is in mid-December with the final deadline in early March, so interviews proceed from late October through April.) Faculty led interviews include introductory remarks followed by general statement-of-purpose and interest questions, and then technical questions in each of the three prerequisite areas. (If a candidate is currently enrolled in one of the prerequisite classes the interviewer can either focus on the other areas or adapt questions to topics already covered in the course.) Faculty then write notes and submit a preliminary recommendation (admit/deny/waitlist). The program director and assistant director then give a final review and arrive at a decision. Small scholarships are often given to help recruit a strong and diverse class. (Average awards are around $4000.) Application season comes to a close with final decisions and offers happening in early to mid-May.

Program admissions data from 2013 - present is summarized in Table 9 below.

Table 9

Year	Submitted Apps	Admitted	Gross Deposits	Melt	Net Deposits by July	Acceptance rate	Enrolled	Yield	Graduated	Graduation Rate
2013	127	33	32	2	30	26%	26	79%	23	88%
2014	241	43	37	2	35	18%	35	81%	33	94%
2015	329	57	53	5	48	17%	46	81%	41	89%
2016	582	101	91	26	71	17%	67	66%	60	90%
2017	688	160	121	25	102	23%	99	62%	80	81%
2018	657	191	18	35	101	29%	95	50%	83	87%
2019*	825	236	134	23	111	29%	101	43%	n/a	n/a

** as of June 2019*

Data Institute

The Data Institute ("DI") was founded in August 2016 with a mission to create industrial and academic partnerships, train highly talented, ethical data scientists, support data science research, build an inclusive community of data scientists and apply data science solutions to social issues. The DI serves as the umbrella organization for data science at USF working with the BSDS and MSDS degree programs as well as all continuing data science education programs. The DI was created in part to handle the various activities that sit outside of the traditional master's degree program: consulting, grants, research and external partnerships. Since its launch nearly three years ago, the DI has grown from four corporate and nonprofit members in year one to 24 in year three (Delta Analytics, Alien Vault, kWh Analytics, Reputation.com, Reddit, Metromile, Schmidt Family Foundation, MTC, Zume, Fair, TruStar, Split, Cogitativo, LeanPlum, NakedPoppy, First Republic Bank, United Healthcare, Ultimate Software, Dome9, ValiMail, Recology, Beam, Bolt Threads, PG&E). The DI is also engaged with foundational support for research initiatives focused on the environment and medical applications of machine learning, a National Science Foundation (NSF)-funded conference attracting researchers from industry and academia and a postdoctoral fellowship program.

The specific areas of support provided to the MSDS program include practicum projects, research support for faculty and continuing education courses for current students and alumni. The DI corporate members provide practicum projects for MSDS students. Running a successful practicum program with industry is a challenge. The DI membership model (with fees ranging from $5,000 to $25,000 to participate) strengthens the engagement and lowers the melt on projects with industry, provides more direct staff support in the recruitment process and provides additional value to companies. Companies join the DI for a number of reasons including access to the talent pipeline of MSDS students as well as discounting and access to continuing education programs in Data Science. The DI also has brought in research focused practicum projects through the Wicklow AI Medical Research Initiative (WAMRI) and through the DI partnership with the Schmidt Family Foundation with projects focused on data science environmental work. Both foundation partnerships have provided faculty research funding, computing resources and scholarship funding to MSDS students. This year alone, eight students participated in WAMRI projects and 3 students participated in Schmidt projects. Students participating in these projects have co-authored papers and their work has been presented at conferences.

In addition to practicum projects, the DI has also provided additional education resources to MSDS students through the continuing education programs. MSDS students have taken Foundations of Deep Learning Using PyTorch (18 MSDS students), Deep Learning Part I (65

MSDS students 2018, 55 MSDS students 2017, 30 MSDS students 2016), Deep Learning Part II (21 MSDS students 2019, 16 MSDS students 2018, 20 MSDS students 2017) and A/B Testing (4 MSDS students, before it was included as a required course in the MSDS curriculum). Given the nature of the cohort based, accelerated program with just one elective course in the final module, the continuing education courses have provided a great opportunity for MSDS students to take more than one 'elective' by allowing them to enroll in Deep Learning or A/B Testing outside of the degree program. MSDS students access these courses for no additional cost while enrolled in the MSDS program.

The DI has also been able to provide extracurricular learning opportunities to MSDS students through consulting projects. This year four current students and three MSDS alumni participated in the ABC News election project to help call the US House of Representatives election. This work resulted in extensive data engineering and modeling building up to the election night. On the night of the election the students were working to ensure their data pipelines and models were correctly functioning to feed the New York City based ABC Decision Desk accurating and early election calls. This project was a resounding success and ABC has asked the Data Institute back to consult on the 2020 election. Addition projects offered to students include participation in Arizona State University survey consulting project. These opportunities provide a chance for the students to apply their skills in a real-world context outside of the nine-month practicum project.

MSDS students were also able to participate in the DI annual conference the past two years. Each year the conference has included 60+ invited speakers from industry and academia. During the October 10-12, 2017 conference, students were able to attend any of the concurrent speaker session on deep learning, networks, experimental design, compressed sensing and machine learning, data science for social good, machine learning and health, and search and optimization. We decided to move to the conference to the Spring semester in large part to better accommodate the MSDS students. By moving the conference to March 10-12, 2019, students were able to attend sessions and present their work at the poster session. Eight students presented work at the conference and we also able to attend speaker session in deep learning, networks, data science for marketing, high dimensional signal processing and machine learning, design of experiments, and technology and data science as the accelerants for sustainable development goals.

Challenges

In this section we outline some of the challenges facing the program.

Resources: As described in detail in the faculty section above, MSDS feels like a program operating with a minimal functioning level of faculty and staff resources. The lack of redundancy, together with institutional and administrative cultures that encourage disciplines to silo on one campus, represent a considerable threat. USF was an early player in the data science education space—a testimony to the foresight of the founding faculty and the nimbleness of administration to quickly roll out the program. The program's quality and success have fostered remarkable growth and a solid network of alumni with impressive pedigree who are well entrenched in the data science ecosystem of the Bay Area (and increasingly beyond). Still, this early adopter advantage is quickly disappearing. The number of competitor programs has exploded of late and new programs continue to arrive on scene. Moreover, new programs arising from colleges of engineering or science threaten the market position of USF as a more

technical master's program in data science. Finally, many of the distinguishing features of the program—one year duration, nine-month practicum, close faculty mentorship, industry engagement—require an unusual level of effort and commitment from faculty. This effort is concentrated on a relatively small number of people with little ability to spread that work out. As a result, faculty burnout is a serious issue. Moreover, the constant administrative demands of the program take time away from equally important faculty responsibilities like academic research and community and professional service.

It's worth noting here that many institutions have announced university-wide initiatives focused on data science and artificial intelligence (AI). A few examples:

Berkeley: https://evcp.berkeley.edu/initiatives/division-data-science-announcement-frequently-asked-questions

Boston University:
http://www.bu.edu/datascience/

Mount Holyoke:
https://www.mtholyoke.edu/media/mhc-destination-data-science

Oregon:
https://datascience.uoregon.edu

Purdue:
https://www.purdue.edu/data-science/

SUNY Binghamton
https://www.binghamton.edu/transdisciplinary-areas-of-excellence/data-science/index.html

Tufts:
https://viceprovost.tufts.edu/data-intensive-study-center/

UC Davis:
https://leadership.ucdavis.edu/initiatives/provost-initiatives/work-group-data-sciences

University of Texas, San Antonio:
https://www.utsa.edu/strategicplan/tactical-initiatives/sds/index.html

While many of these schools already enjoy more resources to draw upon, for example, large graduate departments of mathematics, statistics, and computer science, they continue to invest heavily in the future of data science and data science education. These initiatives span a wide range of approaches and actions, including creation or repurposing of space to facilitate data science collaboration, targeted hiring focus, research and development grants, and a focus on integration of data science throughout the university community in both research and education. There is little indication that USF administration sees data science as an area of particular focus at USF.

Lack of support around career advising, external relations, and alumni engagement.

As discussed above, MSDS is a job-oriented pre-professional program. Outcomes are measured primarily in terms of employment of students. As a result, the program undertakes significant curricular and co-curricular activities in connection with this goal. Examples include:

- Recruiting, vetting, and executing 9-month practicum for each of our 80+ students with industry, government, and non-profit partners.
- Delivery of 15 hours of interview workshops. (For at least five years, the program supported a one-unit course called Interviewing Skills before its faculty voted to move the program to a non-credit workshop model.)
- Ad hoc workshops related to resume and cover letter writing. Program faculty members edit every student's resume and several versions of their cover letters.
- Faculty are deeply involved in advising students about compensation-related negotiations as they seek, and accept, job offers.

At present, these activities are undertaken almost exclusively by the MSDS faculty with no support from either the Office of Career Services on the main campus *or* the SOM-specific career services personnel located at the USF's Downtown Campus. There is some limited release for practicum as well as the resume and interview workshopping, but it fails to reflect the scope of the work done, and represents a significant mismatch between the job that faculty were hired to do and the peripheral but important work necessary in order for the program to succeed.

It is noteworthy that graduate education at USF has grown considerably over the last several decades. USF thirty years ago had virtually no graduate programs. Today, 40% of student credit hours, and one-third of net tuition is generated by graduate programs. Still, the University lacks a coherent strategy regarding student and career services for graduate students. CAS offices for student and career services are located on main campus and are, as far as we can tell, almost entirely undergraduate facing.

Other faculty and staff duties include alumni engagement, external relations, development, grant-writing and administration, and traditional student services (e.g., library, counseling and psychological services, etc.). In many cases, the program's faculty have developed their own creative or entrepreneurial responses to lack of services. For example, the Data Institute was effectively formed to serve the program's external relations needs (and more). But the Data Institute itself requires (and, admittedly, generates) resources. Solutions that require additional uncompensated work by the program's faculty are ultimately not sustainable.

Space

The Master of Science in Data Science program is housed at the USF's Downtown Campus at 101 Howard Street. Despite the fact that the MSDS program resides in the College of Arts and Sciences, the administrators of the downtown campus building report to the School of Management. This arrangement has a long and checkered history, but suffice it to say that there has been an ongoing stream of negative experiences for our students, faculty, and staff.

MSDS was founded as a joint program supported by both the College of Arts and Sciences and the School of Management. In academic year 2014-2015, the then new dean of the School of Management attempted to interfere with program operations, culture, and long-standing practices in a manner perceived by program faculty as non-consultative, non-collegial, and disrespectful. As a result, the program's faculty voted unanimously to recommend to the Provost

that the program be moved solely over to the College of Arts and Sciences for academic administration. The then-current Provost, Jennifer Turpin, accepted this recommendation.

Following this change, there was an attempt by the then dean of the School of Management to retaliate against the program by having it, and its faculty and students, removed from the Downtown Campus. This request was also denied by Provost Turpin, who clarified that the Downtown Campus would become a home for graduate programs from across the University rather than just the graduate programs of the School of Management.

The University has a downtown space committee consisting of dean level and above administration, but the operations and role of that committee are somewhat mysterious. Space proposals are made each year by both CAS and SOM programs and are heard by the committee. Associate deans represent the requests from their affiliated programs. From the perspective of the MSDS program and the DI, there is a strong case to be made for additional space. The most desirable space is on the fifth floor which hosts the only communal student space in the building. We have 80+ students on campus much of the day, and several faculty offices already located on that floor. The fifth floor space is also shared by SOM faculty who also put in requests for space. Not surprisingly, SOM often requests increased space on the fifth floor as well. The unfortunate result of these competing requests is a situation where one program wins and another loses with respect to space. Also, there are old issues that continue to arise in these negotiations. From the perspective of MSDS, however, SOM need on the fifth floor is unclear. Notably, there are unused and empty offices already belonging to SOM on the fifth floor. In light of this situation, and the fact that current MSDS faculty like Wilson and Wang have no dedicated space and are asked to share a hotel office with other faculty, any request for additional space by SOM (especially space currently occupied by MSDS) seems unwarranted. Without question SOM would have their own perspective on issues regarding space.

Currently, the day-to-day operations of the downtown campus is effectively administered by two individuals: Doug Hayward and April McKay. Hayward is an employee of Coldwell Banker Richard Ellis (CBRE), but is attached to the University of San Francisco. He reports up to Facilities / Physical Plant (i.e., Mike London) and is largely concerned with facilities-related issues including servicing the building's non-USF tenants.

The Office Manager for the downtown building reports to Tony Patino, an associate dean of the School of Management. The office manager, therefore, indirectly reports to the same dean who was associated with the program's movement from SOM to the College of Arts and Sciences.

There is a now a well-documented history of poor treatment of the MSDS program. The program's faculty believe that it is past time that the administrative structure of the building change to accommodate the University's vision of the Downtown Campus as a home for graduate programs from across the University.

Examples compiled from MSDS faculty and staff follow.

- Unnecessary and excessive admonishment of our students. Telling them to be quiet and that they may not conduct academic or advisory conversations above a whisper in the fifth-floor agora — because such conversations "distract the front desk."

- SOM personnel routinely fail to remove the adjustable walls in the first-floor classrooms (SFH 154, SFH 155, and SFH 156) to accommodate our program's lectures to its entire

student population (nearly 100 students). Only SOM personnel have the keys required to raise and lower these walls.

- Not granting the program's faculty access to certain shared faculty resources (e.g., hotel offices). This behavior became the argument for the program securing its own dedicated space for faculty to conduct office hours when they travel from the main campus. It also houses certain computational resources.

- Building staff directly support non-academic, non-curricular functions for SOM, but will not provide basic support for data science functions. The program's faculty believe that if staff reported up to a neutral third party, they would be more supportive of non-SOM programming throughout the building.

- Recently, the office manager was pulled away from her responsibilities at the downtown campus to serve as the personal assistant for the School of Management's dean for several weeks. This absence exacerbated a number of building issues and once again demonstrated that building staff support the School of Management and not all building constituents.

- The School of Management has a spatial footprint that cannot be justified by its enrollments or growth trajectory.

- The School of Management and Events Management and Guest Services have established an inadequate calendaring system for the scheduling of common meeting rooms. The current system is driven by email. There is no transparency into room availability, making it difficult to both schedule meetings or have ad hoc collaborations. All information about room reservations must be obtained from personnel reporting up to the School of Management.

- Room scheduling lacks appropriate prioritization. A class of twelve students can book a room and leave a class or event for forty without space.

Another example of a significant challenge facing MSDS in the area of space and support arose in the summer of 2019. The program was informed that the existing practice of scheduling quizzes and exams in the one large room that will accommodate all the students was being ended and that the scheduled bookings for the fall module were being canceled. The program would be forced to try to book the only room that fits its students for assessment activities as an event after census date. As a module based program, this affects our program disproportionately. The accelerated and intense nature of the program means that quizzes and exams are particularly important and a point of particular stress for students. In a seven-week module packed with content, it is impractical to use class time for assessment. While it's understandable that such a policy makes sense in the context of semester long courses with manageable enrollments, we need some flexibility in the context of our module based program with over ninety students. The feasibility of programs like ours rests upon policies that make it easier, and not more difficult, for faculty to do their work and deliver a high quality program.

Summary

The MSDS program has experienced incredible growth from 2012-2019. It has gone from a fledgling program with twelve students to one of the University's largest, welcoming nearly ninety students per year. The success of the program is most vividly captured by its student outcomes: more than ninety percent secure employment within three months of graduation and have a high a median salary that remains constant across demographic and gender breakdowns. The program has an active and collegial group of core faculty who are teacher-scholars dedicated to the ongoing success of the students and the program. The curriculum is carefully planned, rigorous, and current.

Of course, any period of rapid growth comes with challenges. The MSDS program is one of USF's largest graduate programs, and graduate education is an area of particular growth and focus at USF. That said, the University lacks much in the way of infrastructure, support, and resources required to foster and grow strong graduate programs at scale. Examples are detailed above but include lack of adequate student and faculty space, lack of knowledgeable and grad-focused career services (especially for graduate students in tech fields), and lack of support for alumni engagement. Consequently, the responsibilities in these domains either fall through cracks, or fall upon the program's faculty and staff. This only highlights and exacerbates the already inadequate faculty and staff resources available to the program. Moreover, data science education, both graduate and undergraduate, continues to see rapid growth with many new programs appearing each year, making it increasingly difficult (and important) for USF to maintain a competitive advantage and clearly advertise its value proposition.

Course Descriptions

Computation for Analytics (1)

An intense review of Python programming and an introduction to a variety of computational problems. Topics include functions, recursion, loops, list comprehensions, elementary data structures, reading and writing files, image processing, and gradient descent

MSDS 502

Review of Linear Algebra (1)

Topics include matrix operations, linear systems of equations, vector spaces, linear independence, basis and dimension, row/column space, and the rank-nullity theorem; eigenvectors, eigenvalues, and diagonalization of matrices; LU, spectral, and SV

MSDS 504

Review Probability and Stats (1)

A review of random variables, moments, and maximum likelihood estimation; elementary hypothesis tests and con dence intervals; Kolmogorov's axioms, independence, the Law of Total Probability, and Bayes' Theorem; and multivariate distributions, conditional expectation, and Bayesian...

MSDS 593

EDA and Visualization (1)

In this course, students explore sets of data to uncover meaningful associations, detect outliers and anomalies, and formulate hypotheses. Students use R and RStudio to visualize, wrangle, manipulate, and explore data of many types and...

MSDS 601

Linear Regression Analysis (2)

This course is an intensive introduction to linear models, with a focus on both principles and practice. Examples from finance, business, marketing and economics are emphasized. Large data sets are used frequently. Topics include simple and multiple linear regression; weighted, generalized, and...

MSDS 603

Product Analytics (2)

In this course students will develop a minimum viable data product. Using case-studies, students will learn about data-focused companies, their strategies, opportunities and challenges. Some traditional business frameworks will be presented to assist in evaluating strategic data...

MSDS 604

Time Series Analysis (2)

A survey of the theory and application of time series models using R. Tools for model identi cation, estimation, and assessment are developed in depth. Trend and seasonal decomposition models (e.g., Box-Jenkins) are covered, as are smoothing techniques (e.g.,

MSDS 605

Practicum I (1)

The practicum is a data science project sponsored by a company and mentored by a faculty member, allowing students to apply skills alongside industry partners to gain experience, and reconcile mathematical or computational theory with business practice in an apprenticeship style of...

MSDS 610

Communications for Analytics (1)

In this course, students will learn essential concepts related to business communication and, in particular, the communication of technical material both spoken and written. Students will learn how to competently create, organize, and support ideas in their business presentations. They will deliver...

MSDS 621

Intro to Machine Learning (2)

This course focuses on the implementation and application of supervised and unsupervised machine learning algorithms using Python and related libraries. Students learn to properly select features and evaluate model accuracy. Models include at least kNN, naive Bayes, random forests, and...

MSDS 623

Multivariate Statistics (2)

This course trains students in the use of multivariate statistical methods other than multiple linear regression, which is covered in MSDS 601. Applications to finance, social science, and marketing data are emphasized (e.g., dimension reduction for Treasury yield curves and consumer microdata)....

MSDS 625

Practicum II (2)

Students continue to develop skills alongside industry partners and faculty mentors. They gain real-world experience, and reconcile both mathematical and computational theory with business practice in an apprenticeship style of

MSDS 626

Case Studies in Data Science (2)

This course focuses on the application Data Science to solve real problems. It covers a broad selection of key challenges in and methodologies for working with complex data. Topics to be covered include modeling, analysis, visualization, prediction and informed decision making, as well as data security and data privacy. This course is integrative across the core disciplines of Data Science, including databases, statistics, data mining, data visualization, high performance computing, cloud computing, and business intelligence. Students will acquire a working knowledge of data science through hands-on projects and case studies in a variety of business, engineering, social sciences, life sciences domains, among others. Case studies will be drawn from various resources, including data science blogs, as well as recent data challenges from tech companies. Topics will be drawn from recent problems arising in interviews and data challenges.

MSDS 627

Practicum III (2)

Students continue to develop skills alongside industry partners and faculty mentors. They gain real-world experience, and reconcile both mathematical and computational theory with business practice in an apprenticeship style of

MSDS 629

Experiments in Data Science (2)

A survey of statistical methods and best practices surrounding the design and analysis of experiments in the Jeld of data science. A/B tests, A/B/n tests, factorial and fractional factorial designs, response surface methodology and multi-armed bandit experiments will all be...

MSDS 630

Advanced Machine Learning (2)

Students study advanced machine learning algorithms, including boosting, collaborative filtering, support vector machines, expectation maximization for Gaussian mixture models, hidden Markov models, and deep learning neural networks. Teams of students carry out a large-scale real-world...

MSDS 631

Special Topics in Analytics (1 - 2)

Topics will be selected from geographic information systems (GIS), political analytics, sports analytics, supply chain analytics, optimization and simulation, and marketing

MSDS 632

Practicum IV (1)

A continuation of MSAN 627 and conclusion of the practicum program. Students apply skills alongside industry partners and faculty mentors to finish projects and produce quality deliverables, as they fully integrate theoretical analytics coursework with the demands of their practicum…

MSDS 633

Ethics in Data Science (1)

This course introduces ethical and privacy problems in data collection, analysis and evaluation. This course is designed to 1) understand the context of the data science domain focusing on social issues caused by biased data collection and analysis; 2) reflect on these problems, and 3) review and evaluate proposed or possible solutions through readings and hands-on exercises. This class will cover topics including ethical problems in data collection, bias caused by data collection, algorithms, and data evaluation, and their effects on issues of social justice. We will also discuss problems of gamification and addiction of users, and discuss the effects of data science and technology in this context.

MSDS 640

Seminar Series I

Students learn from presentations given by academic researchers, technology executives, practicing data scientists, and business analysts from the Bay Area and beyond. These presentations are open to the

MSDS 641

Seminar Series II

A continuation of MSDS 640. Students continue to participate in weekly presentations and discussions led by local business analysts, data scientists, program alumni, and academic

MSDS 642

Seminar Series III

Students continue to participate in weekly presentations and discussions led by local business analysts, data scientists, program alumni, and academic

MSDS 643

Seminar Series IV

Students continue to participate in weekly presentations, but at this point during the program, the seminar series also provides students with critical networking and job search

MSDS 644

Seminar Series V

A continuation of MSAN 643. Students continue to participate in weekly presentations as the program draws to a close. At this point, the seminar series provides students with critical networking and job search

MSDS 689

Data Structures and Algorithms (1)

This course give students a deeper and more general view of data structures and algorithms. While students have examined a number of data structures, such as binary trees, already, this course provides a much more in-depth study. This depth will benet them greatly in the advanced machine learning course that follows in the next module. This course also tends to address many of the dicult algorithm questions students get during job interviews. The critical data structures covered in this class are: lists, linked lists, trees, graphs, hash tables, and tries. The course also covers a variety of common and useful recursive and non-recursive algorithms, such as searching and sorting.

MSDS 691

Relational Databases (1)

An introduction to relational databases focusing on learning SQL with the Postgres database. Topics include schemas (tables in various normal forms), indexes, query efficiency, server-specific navigation functions, and queries with grouping, ordering, sorting, collapsing, and...

MSDS 692

Data Acquisition (2)

This lab-heavy Python class teaches students how to collect, merge, and clean data from multiple sources and organize it into appropriate data structures. Topics include XML, JSON, HTML, REST APIs, scraping data from websites, and using Selenium to extract data from JavaScript-based...

MSDS 694

Distributed Computing (1)

Students learn the MapReduce technique of distributed computing. The fundamental principles are first learned with the Python multiprocessing library, in which students build their own concurrent MapReduce framework. Considerable time is spent exploring practical application of mapping and...

MSDS 697

Distributed Data Systems (2)

Students study key-value store through NoSQL with a focus on using MongoDB (including, possibly, pymongo, the Python Mongo API). Applications are used to motivate a disciplined approach to database programming with MongoDB, including the construction of

MSDS 699

Machine Learning Laboratory (1)

This course is a lab course associated with the MSDS machine learning course (MSDS621). MSDS621 focuses on the implementation of models, whereas, in this lab, students learn to use pre-existing implementations of these machine learning models from the scikit-learn machine learning library: https://scikit-learn.org. Machine learning is about much more than the models themselves. The ecosystem of tools and libraries surrounding these models are just as important as the models. For example, effective use of pandas dataframes, https://pandas.pydata.org is a prerequisite for developing machine learning models. There are also a number of techniques that are important to learn in order to use models eec-tively. For example, knowing how to properly assess the performance of a model is critical. Models must be tested using validation and test sets, not the training data used to train the model itself. (Time-based data sets make such assessment even more complicated.) Another critical technique is so-called feature engineering, whereby we provide the richest and most predictive features possible to help the model capture the relationship between predictors and target variables. Students will learn the key libraries and basic processes necessary to apply machine learning.