

Competitive Analysis

A look at top-ranked institutions' data science programs



An Introduction

This document contains accumulated information regarding data science programs from top research institutions from around the country. The universities studied include UC Berkeley, Carnegie Mellon, Harvard, MIT, Stanford, and Yale. For each college, there is detailed information about the school's funding, faculty diversity, course structure, and research as it pertains specifically to each institution's data science environments. After carefully studying each school's involvement in data science, they were given a score out of ten for each of the five following categories:

Interdisciplinary Faculty: How diverse were there faculty in data science programs?

CS/Stats Departments: How well do these schools rank individually in terms of statistics and computer science.

Industry Partners/Funding: Do these school attract funding for their initiatives?

Community: Are there student organizations and events on campus creating a community for data enthusiasts?

Education: Does this school offer multiple degree paths for students to benefit from?

After evaluating institutions on these five categories, the college was given a total score out of 100 as an indication of how well each school fares in data science education in relation with the others.

BERKELEY

About the Initiative



"Founded in 2013, the Berkeley Institute for Data Science (BIDS) is a central hub of research and education at UC Berkeley designed to facilitate and nurture data-intensive science. People are at the heart of BIDS. We are building a community centered on a cohort of talented data science fellows and senior fellows who are representative of the world-class researchers from across campus and are leading the data science revolution within their disciplines."



Industrial Funding



BIDS received its first financial sponsorship from the Moore and Sloan Foundations with a 5-year \$12.5 million grant and is one of 3 university Data Science Environment Partnerships.



ALFRED P. SLOAN FOUNDATION

Industry Partner Program

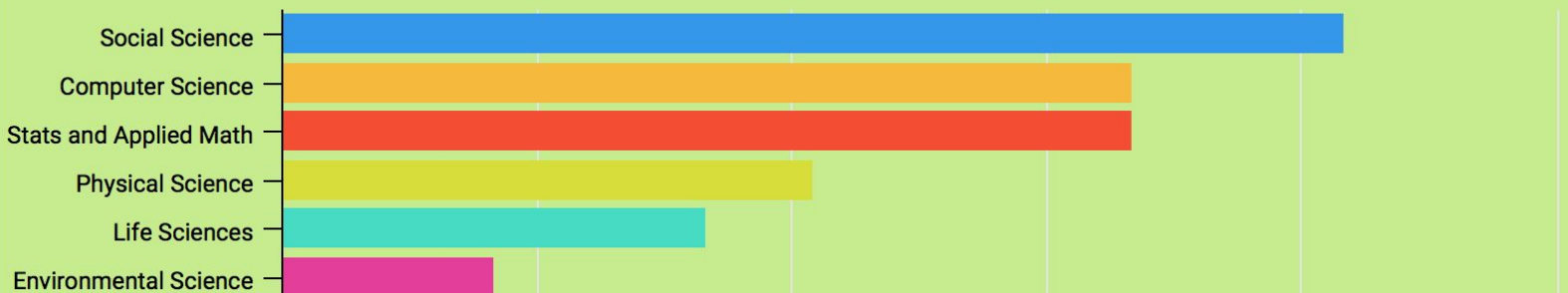
The BIDS Industry Partner program seeks to connect the world-class data science research teams at UC Berkeley with leaders in industry engaged in extracting value through analytics on vast, subtle, and often-imperfect data streams.



Faculty and Staff Diversity



The data science initiative at UC Berkeley encompasses BIDS (the research center for data science), Data Science Education Program (DSEP), and the diverse courses taught by faculty from various disciplines. Like most other schools, the initiative is comprised of CS and Stats backgrounds, but UC Berkeley is uniquely applying a predominant focus of data science applications in the social sciences.



Course Structure



UC Berkeley is currently in the works of establishing data science degree programs for undergraduates, including a data science major and minor. Berkeley is also paving the way for a School of Computation for interdisciplinary data science programs. For graduate students there is an online masters program in data science.

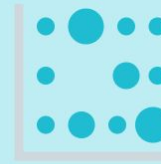
Currently Establishing



Masters in Data Science (MIDS) offered by School of Information



Undergraduate Major in Data Science



Undergraduate Minor in Data Science



School of Computation and Data Science

Research Projects



Listed below are some of the several labs at UC Berkeley that focus on data science related projects in fields like humanities, biological sciences, computer science, geomatics, and more.



AMPLab: Algorithms, Machines, People



BRC: Berkeley Research Computing



Scalable Data Management, Analysis, and Visualization



GIF: GeoSpatial Innovation Facility

Succinct on Apache Spark: improve search

Prosopography: NLP to study Babylonian society

FastBit: efficient search, helping optimize Gene Context Analysis

Holos: the Berkeley Ecoinformatics Engine

Rankings



Interdisciplinary Research/Faculty: Berkeley offers a wide array of both faculty and research in data science. Strength factors include specific labs dedicated to data science research.

With Respect to CS/Stats Department: Berkeley ranks first in computer science and second in statistics.

Industry Partners/Funding: Despite having strong initiatives in data science, Berkeley lacks industry sponsors in comparison to other schools.

Community: Berkeley fosters a strong community around data science through student clubs like Data Science Society, Berkeley Institute of Data Science, Machine Learning at Berkeley, and more.

Education: Berkeley has a masters program currently, but does offer promise in the future regarding more programs for undergraduates.

8

9

5

8

7

Total Score
74

CARNEGIE MELLON

About the Initiative



"Carnegie Mellon University's programs in Data Science are designed to train students to become tomorrow's leaders in this rapidly growing area. Through a unique combination of interdisciplinary coursework and cutting-edge research, the programs will enable them to apply techniques and tools of data science to applications drawing on appropriate and relevant concepts and models from the engineering, natural or social sciences."



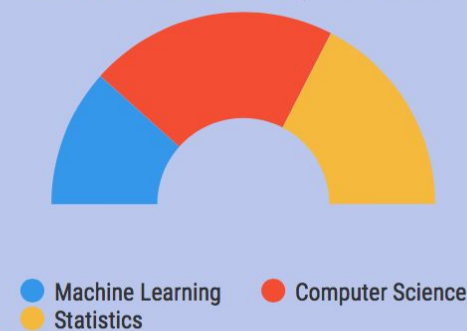
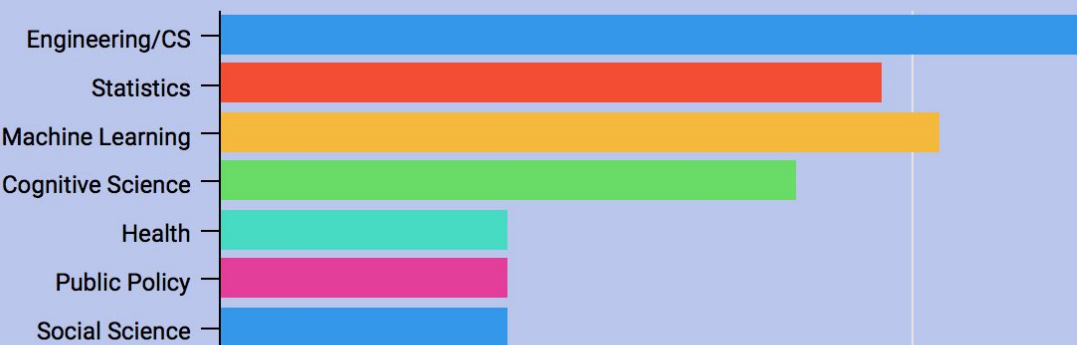
Faculty Diversity



Carnegie Mellon provides students with a variety of diverse faculty, with computer science, statistics, and machine learning faculty leading the way.

Fields Corresponding to Faculty at Carnegie Mellon

Main depts. constituting data science education at Carnegie Mellon



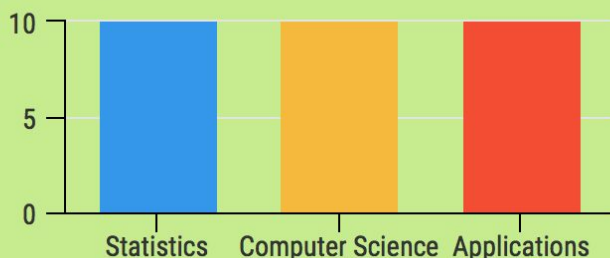
Masters Program Structure



The masters program for data science at Carnegie Mellon University offers a core foundation in data science, but also allows graduates to specialize in applied fields such as policy, business, and education.

Structure of Data Science Program

Carnegie Mellon places equal importance among these three fields.



Examples of Application Pathways

- Business Analytics
- Information, Systems Management
- Language Technologies
- Education Technology
- Public Policy and Management
- Intelligent Information Systems
- Business Administration



Course Structure



Carnegie mainly offers graduate degrees in data science, and these degrees are awarded from the a variety of department such as statistics, computer science, and machine learning. Though there is no specified data science track in undergraduate education, they do offer a a major and minor through dept. of machine learning.



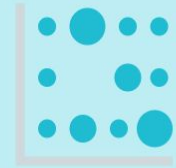
BS in Statistics and Machine Learning
(Dept. of Machine Learning)



Masters in Computational Data Science
(Dept. of CS)



Masters in Statistical Practice
(Dept. of Statistics)



Masters in Machine Learning
(Dept. of Machine Learning)

Research Projects



Projects administered in Carnegie Mellon place strong emphasis on interdisciplinary research. Though there remains a health focus, Carnegie Mellon's projects explores a variety of fields including sociology, cognitive science, natural language processing and more.



Embedded Assessment of Cognitive Decline



Monitoring Human Performance Using Accelerometers



First-Person Vision Usability Study



Personal Health Coach



Recording Observations Elderly of Daily Living

Rankings



Interdisciplinary Research/Faculty: Carnegie Mellon offers a wide array of programs, research, and faculty. This enables a diverse data science environment at the university.

8

With Respect to CS/Stats Department: Carnegie Mellon ranks first in computer science, but only ninth in statistics.

7

Industry Partners/Funding: Carnegie Mellon offers very little information on their industry partners and funding for data science.

5

Community: Despite having a strong programs, Carnegie Mellon has very few student organizations or events corresponding to data science.

6

Education: Carnegie Mellon offers an array of data science degrees through graduate programs, but very limited are offered in undergraduate education.

8

Total Score

68

MIT

About the Initiative



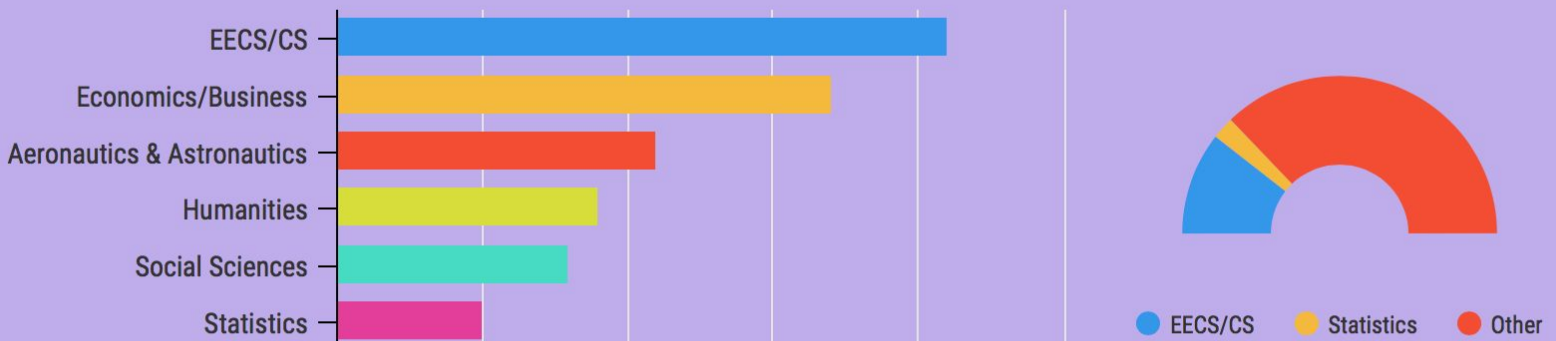
MIT's Institute for Data, Systems, and Society

"Spanning all five schools at MIT, IDSS embraces the collision and synthesis of ideas and methods from analytical disciplines including statistics, data science, information theory and inference, systems and control theory, optimization, economics, human and social behavior, and network science. "

Faculty Diversity



While many of the faculty in IDSS come from an EECS/CS and Economics/Business background, there is a diverse population of members from fields in political science, media arts and studies, computational linguistics, civil and environmental engineering, sociology, anthropology, operations research, and more.



Industry



IDSS and its industry connections work together to realize IDSS's mission and goals—namely, to gather and analyze large sets of data, develop new models to effectively interpret, and ultimately, to use this data to inform solutions to real-world problems.

Founding Members



THOMSON REUTERS

WORLDQUANT.

Booz | Allen | Hamilton

Benefits of Partnership

Involvement in IDSS's External Advisory Board

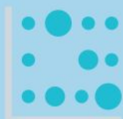
Access to IDSS students to work for selected challenges

Access to faculty for project consultations

Course Structure



IDSS offers a range of degrees related to the field of data science in partnership with institutes such as the Statistics and Data Science Center, which offers an undergraduate minor in statistics and data science. IDSS students learn from a variety of disciplines from information and decision systems, data science, statistics, and social sciences with application domain areas to address societal challenges.



Undergrad Minor
Statistics and Data Science

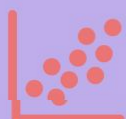


Doctoral Program
Social and Engineering Systems



Technology and Policy
Program (TPP)

Research



Statistical and
Computational Tradeoffs



Causal Inference and Applications
to Learning Gene Regulatory
Networks



Online Learning



Nonparametric Bayesian
Statistics

Industry Expertise



Energy Systems



Finance



Healthcare



Social Networks



Urban Systems

Rankings



Interdisciplinary Research/Faculty: IDSS boasts an impressive diverse faculty from a wide range of disciplines.

With Respect to CS/Stats Department: MIT ranks #1 in Computer Science and Electrical Engineering, and while it does not have a statistics program, it is ranked #6 in industrial/manufacturing/systems engineering

Industry Partners/Funding: IDSS few industry partners focus less on funding and more on research and management

Community: Amongst MIT's large number of student organizations, there is a strong presence of clubs related to data science

Education: MIT offers a diverse range of degrees and programs either through IDSS or Sloan Business Analytics major

9

6

5

7

8

Total
Score
70

STANFORD

About the Initiative



"Stanford Data Science Initiative (SDSI) consists of data science **research**, shared data and computing **infrastructure**, shared **tools** and **techniques**, **industrial links**, and **education**. SDSI has strong ties to groups across Stanford University such as medicine, computational social science, biology, energy, and theory."

Industrial Funding



One of the key strengths of SDSI is its ability to attract industrial sponsors. SDSI has two groups -- **Founding** and **Regular** members. To be classified as one of the two, there are minimum donation requirements and benefits corresponding to the group.

Founding Members



\$100,000

Starting Donation



\$500,000

Regular Members



Benefits

- research agenda
- an office on campus
- a faculty liaison
- custom workshops at Stanford
- faculty and student visits to companies.

- research retreats
- conferences and seminars
- informal interactions.

Faculty Diversity

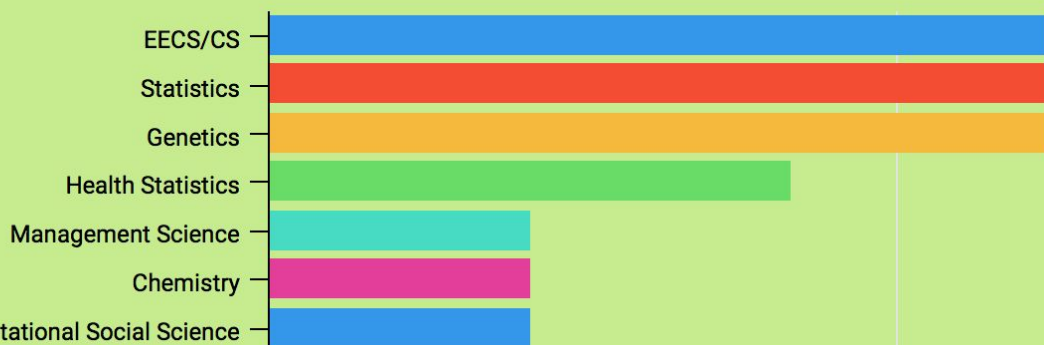


Though the faculty involved in the SDSI come from diverse fields including computer science, statistics, genetics, chemistry, and social science, Stanford's emphasis remains on health-related fields.

Prominence of Health Faculty



● Other
● EECS/Stats
● Health-Related Faculty



Course Structure



Stanford mainly offers graduate degrees in data science, and these degrees are awarded from the department of statistics and institute for computational & mathematical engineering.



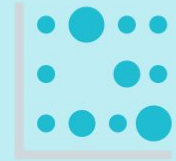
Graduate Certification in Data Mining and Applications



Masters in Computational Mathematics with Data Science Track



Masters in Statistics with a Data Science Path



Phd in Computational Mathematics with Data Science Track

Research Projects



Projects administered under SDSI place strong emphasis on interdisciplinary research. Though there remains a health focus, Stanford's projects explore a variety of fields including sociology, genetics, natural language processing and more.



Mapping the "Social Genome"



DeepDive—a High-Performance Inference and Learning Engine



Secure Analytics on the Internet of Things



Data Science for Personalized Medicine



Data Commons - Repository of data for research at Stanford using cloud computing

Rankings



Interdisciplinary Research/Faculty: Though SDSI contains a diverse array of projects and faculty, Stanford remains best for medicine-related research.

8

With Respect to CS/Stats Department: Stanford ranks first in both computer science and statistics.

9

Industry Partners/Funding: SDSI places strong emphasis on attracting industry funding.

9

Community: Despite having a strong initiative, Stanford doesn't have too many student organizations or events corresponding to data science.

6

Education: Stanford offers an array of data science degrees through graduate programs, but none are offered in undergraduate education.

7

Total Score
78



Data Science at Other Universities



The following universities do not have well-formed data science programs, yet it is still important to consider them in the competitive analysis. Because data science is a new field, the inclusion of schools with an emerging data science programs signify that most top-rank institutions are beginning to incorporate data science, even if it's not to the extent of the other schools in this analysis.



HARVARD UNIVERSITY

In addition to its certificate, Harvard's Statistics program, ranked #7 in the country, has a strong data science track. The university also recently created CS 109, which is open to all undergraduates, graduates, and Harvard Extension students.



CS 109 "Data Science"
Harvard School of Engineering and
Applied Sciences



Certificate in Data Science
Harvard Extension




Statistics
Undergraduate Major
Data Science Track



Yale University

While Yale does not have any notable undergraduate or graduate academic program in data science, the institution's various schools have collaboratively developed interdisciplinary research programs and centers that leverage data science in their respective fields.

Y | CRC
-Center for Research
Computation-

 Yale Center for
Analytical Sciences
-School of Public Health-

Epidemiology
Material Science Agriculture
Biostatistics Computational Psychiatry
Evolution Social Science
Genetics Public Health Neurobiology
Health Economics and Policy Astronomy
Molecular Biology Health Services

YALE
DATA-DRIVEN



Yale University's
School of Forestry



Yale-NUS College
of Singapore

Data-Driven Yale uses innovative data analytics to distill signals from large-scale and unconventional datasets and develop policy solutions to contemporary environmental problems.

Conclusion

After studying the data science institutions at the different universities, it's important to notice that each school incorporates a unique dimension of data science into their respective programs:

UC Berkeley interweaves aspects of social sciences into their program through the form of course projects in core data science classes and much of the research they engage in.

Stanford University places a strong emphasis on medicine and health through their research projects, and this speaks to the university's top-ranked medical program on campus.

MIT's data science programs offers a business perspective and is well-suited for students who have a strong interest in the intersection between business and data science.

As **Carnegie Mellon** offers a specific degree in machine learning, its program is most suited for those seeking to specialize in that area of data science.

Although it is relevant to note the distinctions between data science offerings, we found that each school offers a unique approach to data science based on the culture of their institution. The scores only provide a partial picture of how well established the programs are -- each of them has something distinct to offer in data science. We look forward to seeing their progress in the years to come!

DSEP Mapping Research Team



BERKELEY INSTITUTE
FOR DATA SCIENCE