Applied Data Science

The Graduate Certificate in Applied Data Science, offered by the UC Berkeley School of Information, introduces the tools, methods, and conceptual approaches used to support modern data analysis and decision-making in professional and applied research settings. It exposes students to the challenges of working with data (e.g., asking a good question, inference and causality, decision-making) as well as to the new tools and techniques for data analytics (machine learning, natural language processing, and more).

The certificate is particularly designed to meet the needs of the graduate students in Berkeley's professional schools — both professional master's students and doctoral students — as well as graduate students in the social sciences and the arts & humanities.

The need for expertise in data analytics continues to grow in all organizations and disciplines. Graduate students in every field are now working with data from new sources: websites, electronic medical records, transaction records, sensor networks, smart phones, and digitized records and documents. The analytical tools and methods traditionally used to derive insights from structured and well-curated data sets (census, surveys, and administrative data) are not sufficient for this new, unstructured and often user-generated data.

The Graduate Certificate in Applied Data Science provides hands-on practice working with unstructured and user-generated data to identify new ways to inform decision-making. The curriculum educates professionals and scholars to be intelligent consumers of data science techniques in a variety of domains, with a foundation of skills for applying these techniques in their own domains.

Any UC Berkeley graduate student who meets the following prerequisites is eligible to pursue the certificate.

- be registered and enrolled in a graduate degree at UC Berkeley,
- be in good academic standing,
- meet course and subject matter prerequisites for courses taken in the certificate program, including Python programming and basic statistics knowledge.

The certificate completion application should be submitted on the School of Information website during the second half of the semester in which the student completes the certificate requirement. All three courses must be either completed or currently in progress. Applications are accepted three times a year, in the second half of the fall, spring or summer semesters.

Prerequisites

Applicants must:

- Be registered and enrolled in a graduate degree program at UC Berkeley
- Be in good academic standing
- Meet course and subject matter prerequisites for courses taken in the certificate program, typically including Python programming and basic statistics knowledge.

Certificate Requirements

The certificate requires three 3-unit courses, taken from the following approved lists:

1. An introductory data science class
2. A course in analytical methods and techniques of data science
3. An additional elective: either a domain-specific data science course or a second methods course.

Courses should be taken for a letter grade and must be completed with a grade of B or higher. At least one of these courses must be a course offered outside the student's graduate program.

1. INtroductory data science course

One of the following:

INFO 201	Research Design and Applications for Data and Analysis [1]	3
DATASCI 201	Research Design and Applications for Data and Analysis (MIDS and MICS students only)	3

2. Analytical Methods and Techniques of Data Science

Students must take at least one course from this list:

BIO ENG 245	Introduction to Machine Learning for Computational Biology	4
COMPSCI C200A	Principles and Techniques of Data Science	4
COMPSCI C281A	Statistical Learning Theory	3
COMPSCI 289A	Introduction to Machine Learning	4
CYBER 207	Applied Machine Learning for Cybersecurity (MIDS and MICS students only)	3
DATA C200	Principles and Techniques of Data Science	4
DATASCI 207	Applied Machine Learning (MIDS and MICS students only)	3
EDUC 244	Data Mining and Analytics	3
INFO 251	Applied Machine Learning	4
INFO 258	Data Engineering	4
INFO 271B	Quantitative Research Methods for Information Systems and Management	3
PB HLTH 241	Intermediate Biostatistics for Public Health	4
PB HLTH W241	Course Not Available	4
PSYCH 208	Methods in Computational Modeling for Cognitive Science	3
SOCIOL 273L	Computational Social Science	3
STAT C200C	Principles and Techniques of Data Science	4
STAT C241A	Statistical Learning Theory	3

3. Electives

Students must take one domain-specific data science course from the following list or a second methods course from the list in Section 2 above:

A,RESEC 213	Applied Econometrics	4
CIV ENG 263N	Scalable Spatial Analytics	3
CIV ENG C263H	Human Mobility and Network Science	3
COMPSCI C267	Applications of Parallel Computers	3-4
COMPSCI 286A	Course Not Available	4

COMPSCI C281B	Advanced Topics in Learning and Decision Making	3
COMPSCI 288	Natural Language Processing	4
CY PLAN 204C	Analytic and Research Methods for Planners: Introduction to GIS and City Planning	4
CY PLAN 255	Urban Informatics and Visualization	3
CY PLAN 257	Data Science for Human Mobility and Socio-technical Systems	4
CY PLAN C257H	Human Mobility and Network Science	3
DEVP 229	Quantitative Methods and Impact Evaluation	3
DATASCI 209	Data Visualization (MIDS and MICS students only)	3
DATASCI 241	Experiments and Causal Inference (MIDS and MICS students only)	3
DATASCI 266	Natural Language Processing with Deep Learning (MIDS and MICS students only)	3
EDUC 275B	Data Analysis in Educational Research II	4
EDUC 275G	Hierarchical and Longitudinal Modeling	5
EDUC 276E	Research Design and Methods for Program and Policy Evaluation	3
EECS 227AT	Optimization Models in Engineering	4
EL ENG 227BT	Convex Optimization	4
EL ENG C227C	Convex Optimization and Approximation	3
EL ENG C227T	Introduction to Convex Optimization	4
ENGIN C233	Applications of Parallel Computers	3-4
ESPM 215	Hierarchical Statistical Modeling in Environmental Science	2
ESPM 288	Reproducible and Collaborative Data Science	3
EWMBA 247	Topics in Operations and Information Technology Management (Fall 2022 & Fall 2023, topic “Descriptive and Predictive Data Mining” only)	2
EWMBA 263	Marketing Analytics	3
GEOG 249	Spatiotemporal Data Analysis in the Climate Sciences	3
GEOG 279	Statistics and Multivariate Data Analysis for Research	3
GEOG 282	Geographic Information Systems: Applications in Geographical Research	4
GEOG 285	Topics in Earth System Remote Sensing	3
IND ENG C227A	Introduction to Convex Optimization	4
IND ENG C227B	Convex Optimization and Approximation	3
IND ENG 242A	Machine Learning and Data Analytics	4
IND ENG 262A	Mathematical Programming I	4
IND ENG 262B	Mathematical Programming II	3
IND ENG 264	Computational Optimization	3
IND ENG 265	Learning and Optimization	3
IND ENG 266	Network Flows and Graphs	3
IND ENG 269	Integer Programming and Combinatorial Optimization	3
INFO 213	Introduction to User Experience Design	4
INFO 241	Experiments and Causal Inference	3
INFO 247	Information Visualization and Presentation	4
INFO 256	Applied Natural Language Processing	3
INFO 259	Natural Language Processing	4
INFO 288	Big Data and Development	3
INFO 290T	Special Topics in Technology (“Biosensory Computing” topic only)	2-4
JOURN 221	Introduction to Data Visualization	3
LD ARCH 289	Applied Remote Sensing	3
LINGUIS 252	COMPUTATIONAL LINGUISTICS	3
MAT SCI 215	Computational Materials Science	3
MBA 247	Topics in Operations and Information Technology Management (Fall 2022 & Fall 2023, topic “Descriptive and Predictive Data Mining” only)	2
MBA 263	Marketing Analytics	3
MBA 296	Special Topics in Business Administration (Fall 2019 & 2020, section 7B, & Spring 2023, Section 8: “Data Science Applications in Finance and Accounting” only)	2
MEC ENG 249	Machine Learning Tools for Modeling Energy Transport and Conversion Processes	3
MFE 230P	Financial Data Science	2
PB HLTH 231A	Analytic Methods for Health Policy and Management	3
PB HLTH C240A	Introduction to Modern Biostatistical Theory and Practice	4
PB HLTH C240B	Biostatistical Methods: Survival Analysis and Causality	4
PB HLTH C240C	Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine	4
PB HLTH C240D	Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine II	4
PB HLTH C242C	Longitudinal Data Analysis	4
PB HLTH 244	Big Data: A Public Health Perspective	3
PB HLTH W251B	Course Not Available	2
PB HLTH 251C	Course Not Available	2
PB HLTH 252	Epidemiological Analysis	4
PB HLTH W252	Course Not Available	4
PHYSICS 288	Bayesian Data Analysis and Machine Learning for Physical Sciences	4
POL SCI C236A	The Statistics of Causal Inference in the Social Science	4
POL SCI C236B	Quantitative Methodology in the Social Sciences Seminar	4
POL SCI 231B	Quantitative Analysis in Political Research	4
POL SCI 239T	An Introduction to Computational Tools and Techniques for Social Science Research	4
PSYCH 206	Structural Equation Modeling	3
PSYCH 207	Person-Specific Data Analysis	3
PUB POL 249	Statistics for Program Evaluation	4
PUB POL 275	Spatial Data and Analysis	4
PUB POL 279	Research Design and Data Collection for Public Policy Analysis	3
PUB POL 288	Risk and Optimization Models for Policy	4
PUB POL 290	Special Topics in Public Policy (“Data Science for Public Policy” or “Quantitative Methods and Evaluation” topics only)	1-4
SOCIOL C271D	Quantitative/Statistical Research Methods in Social Sciences	3
SOCIOL 273L	Computational Social Science	3

SOCIOL 273M	Computational Social Science	3
STAT 215A	Applied Statistics and Machine Learning	4
STAT 215B	Statistical Models: Theory and Application	4
STAT 238	Bayesian Statistics	3
STAT C239A	The Statistics of Causal Inference in the Social Science	4
STAT C239B	Quantitative Methodology in the Social Sciences Seminar	4
STAT C241B	Advanced Topics in Learning and Decision Making	3
STAT 243	Introduction to Statistical Computing	4
STAT 244	Computing for Statistics and Data Science with Julia	4
STAT C245A	Introduction to Modern Biostatistical Theory and Practice	4
STAT C245B	Biostatistical Methods: Survival Analysis and Causality	4
STAT C245C	Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine	4
STAT C245D	Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine II	4
STAT C247C	Longitudinal Data Analysis	4
STAT 248	Analysis of Time Series	4
STAT 256	Causal Inference	4
STAT 259	Reproducible and Collaborative Statistical Data Science	4
STAT C261	Quantitative/Statistical Research Methods in Social Sciences	3
VIS SCI 265	Neural Computation	3