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, data mining, 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 in good standing may apply. To apply, students should submit the following materials on the School of Information website (https://www.ischool.berkeley.edu/programs/data-science-certificate/):

- a letter of intent,
- a proposed study plan,
- a description of their Python programming and statistics competencies,
- their curriculum vitae or resume,
- their Berkeley course transcript.

Applications are accepted twice a year, in the middle of the fall and spring semesters.

Students may apply at any time during their UC Berkeley graduate career, either before or after taking courses that would count toward the certificate.

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 an INFO course offered by the School of Information.

1. Introductory data science course

One of the following:

- INFO 201  Research Design and Applications for Data and Analysis
- DATASCI 201  Research Design and Applications for Data and Analysis (MIDS and MICS students only)

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
- COMPSCSI C200A  Principles and Techniques of Data Science
- COMPSCSI C281A  Statistical Learning Theory
- COMPSCSI 289A  Introduction to Machine Learning
- CYBER 207  Applied Machine Learning for Cybersecurity (MIDS and MICS students only)
- DATA C200  Principles and Techniques of Data Science
- DATASCI 207  Applied Machine Learning (MIDS and MICS students only)
- EDUC 244  Data Mining and Analytics
- INFO 251  Applied Machine Learning
- INFO 258  Data Engineering
- INFO 271B  Quantitative Research Methods for Information Systems and Management
- STAT C200C  Principles and Techniques of Data Science
- PB HLTH 241  Intermediate Biostatistics for Public Health
- PB HLTH W241  Intermediate Biostatistics for Public Health
- PSYCH 208  Methods in Computational Modeling for Cognitive Science
- SOCIOL 273L  Computational Social Science
- STAT C200C  Principles and Techniques of Data Science
- STAT C241A  Statistical Learning Theory

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:

- CIV ENG 263N  Scalable Spatial Analytics
- COMPSCSI C267  Applications of Parallel Computers
- COMPSCSI 286A  Introduction to Database Systems
- COMPSCSI C281B  Advanced Topics in Learning and Decision Making
<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 248</td>
<td>Analysis of Time Series</td>
<td>4</td>
</tr>
<tr>
<td>STAT 256</td>
<td>Causal Inference</td>
<td>4</td>
</tr>
<tr>
<td>STAT 259</td>
<td>Reproducible and Collaborative Statistical Data Science</td>
<td>4</td>
</tr>
<tr>
<td>STAT C261</td>
<td>Quantitative/Statistical Research Methods in Social Sciences</td>
<td>3</td>
</tr>
<tr>
<td>VIS SCI 265</td>
<td>Neural Computation</td>
<td>3</td>
</tr>
</tbody>
</table>