DATA SCIENCE (MS)

Department Website (http://cds.nyu.edu)

**NYSED:** 35896 **HEGIS:** 1702.00 **CIP:** 30.7001

**Program Description**

The Master of Science in Data Science is a highly-selective program for students with a strong background in mathematics, computer science, and applied statistics. The degree focuses on the development of new methods for data science.

Our networked world is generating a deluge of data that no human, or group of humans, can process fast enough. This data deluge has the potential to transform the way business, government, science, and healthcare are carried out. But too few possess the skills needed to use automated analytical tools and cut through the noise to create knowledge from big data.

A new discipline has emerged to address the need for professionals and researchers to deal with the “data tidal wave.” Its objective is to provide the underlying theory and methods of the data revolution. This emergent discipline is known by several names. We call it “data science,” and we have created the world’s first MS degree program devoted to it.

The curriculum is 36 credits, and offers two ways to structure the graduate program that gives students the opportunity to pursue a specialization through the Industry Concentration or tracks.

**Admissions**

All applicants to the Graduate School of Arts and Science (GSAS) are required to submit the general application requirements (https://gsas.nyu.edu/nyu-as/gsas/admissions/arc.html), which include:

- Academic Transcripts (https://gsas.nyu.edu/nyu-as/gsas/admissions/arc/academic-transcripts.html)
- Test Scores (https://gsas.nyu.edu/nyu-as/gsas/admissions/arc/test-scores.html) (if required)
- Applicant Statements (https://gsas.nyu.edu/nyu-as/gsas/admissions/arc/statements.html)
- Résumé or Curriculum Vitae
- Letters of Recommendation (https://gsas.nyu.edu/nyu-as/gsas/admissions/arc/letters-of-recommendation.html), and
- A non-refundable application fee (https://gsas.nyu.edu/admissions/arc.html#fee).

Admission to the Master of Science in Data Science requires substantial but specific mathematical competencies, typical of a major in mathematics, statistics, engineering, physics, theoretical economics, and computer science with sufficient mathematical training. In addition, applicants should have some training in programming and basic computer science. Preference is given to applicants with prior exposure to machine learning, computational statistics, data mining, large-scale scientific computing, operations research (either in an academic or professional context), as well as to applicants with significantly more mathematical and/or computer science training than the minimum requirements listed above.

See Data Science (https://gsas.nyu.edu/admissions/arc/programs/data-science.html) for admission requirements and instructions specific to this program.

**Program Requirements**

The program requires the completion of 36 credits, and offers a concentration in Industry.

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS-GA 1001</td>
<td>Introduction to Data Science</td>
<td>3</td>
</tr>
<tr>
<td>DS-GA 1002</td>
<td>Probability and Statistics for Data Science</td>
<td>3</td>
</tr>
<tr>
<td>DS-GA 1003</td>
<td>Machine Learning</td>
<td>3</td>
</tr>
<tr>
<td>DS-GA 1004</td>
<td>Big Data</td>
<td>3</td>
</tr>
<tr>
<td>DS-GA 1006</td>
<td>Capstone Project and Presentation</td>
<td>3</td>
</tr>
</tbody>
</table>

**Electives**

Data Science Elective | 3

Select one of the following:

- DS-GA 1005 | Inference and Representation |
- DS-GA 1008 | Deep Learning |
- DS-GA 1011 | Natural Language Processing with Representation Learning |
- DS-GA 1012 | Natural Language Understanding and Computational Semantics |
- DS-GA 1013 | Mathematical Tools for Data Science |
- DS-GA 1014 | Optimization and Computational Linear Algebra |
- DS-GA 1015 | Text as Data |
- DS-GA 1016 | Computational Cognitive Modeling |
- DS-GA 1017 | Responsible Data Science |
- DS-GA 1018 | Probabilistic Time Series Analysis |

**Other Elective Credits** | 18

**Total Credits** | 36

**Industry Concentration**

Students have the opportunity to pursue a specialization through the Industry Concentration. This concentration is specifically targeted to respond to the needs and inputs from companies and allows MS in Data Science students to apply the knowledge and skills obtained in their coursework to industry during the degree program. It requires more industry-targeted coursework and a Practical Training experience. Students in this concentration will be required to take the following courses for the degree as a part of the 36 credit requirement:

1. DS-GA 1009 Practical Training for Data Science within the first year of the program (3 credits in fall, spring, or summer), and
2. Two electives within the Big Data or Natural Language Processing subject areas (6 credits).

   a. The courses below fall within the Big Data subject area. This list is approved and reviewed annually.
      i. DS-GA 1012 | Natural Language Understanding and Computational Semantics |
      ii. CS-GY 6313 | Information Visualization |
      iii. CS-GY 6323 | Large-Scale Visual Analytics |
      iv. CS-GY 6083 | Principles of Database Systems |
      v. DS-GA / CSCI-GA 2433 | Database Systems |
      vi. CS-GY 6093 | Advanced Database Systems |
vii. CSCI-GA 2434 Advanced Database Systems
viii. CSCI-GA 2436 Realtime and Big Data Analytics
ix. CSCI-GA 2437 Big Data Application Development
x. CSCI-GA 3033 Special Topics in Computer Science: Cloud and Machine Learning
xi. CSCI-GA 3033 Special Topics in Computer Science: Introduction to Deep Learning Systems
xii. INTG1-GC 1025 Database Management & Modeling
xiii. TECH-GB 2350 Robo Advisors & Systematic Trading
xiv. MATH-GA 2047 Trends in Financial Data Science

b. The courses below fall within the Natural Language Processing subject area. This list is approved and reviewed annually.
   i. DS-GA 1011 Natural Language Processing with Representation Learning
   ii. DS-GA 1012 Natural Language Understanding and Computational Semantics
   iii. CSCI-GA 3033 Statistical NLP
   iv. DS-GA 1005 Inference and Representation
   v. DS-GA 1008 / CSCI-GA 2572 Deep Learning
   vi. DS-GA 1015 Text as Data
   vii. CSCI-GA 2590 Natural Language Processing
   viii. CSCI-GA 3033 Special Topics in Computer Science: Learning with Large Language and Vision Models

All other requirements remain the same.

**Capstone (DS-GA 1006)**

One of the key features of the MS in Data Science curriculum is a capstone project that makes the theoretical knowledge gained in the program operational in realistic settings. During the project, students go through the entire process of solving a real-world problem; from collecting and processing real-world data, to designing the best method to solve the problem, and finally, to implementing a solution. The problems and datasets come from real-world settings identical to what might be encountered in industry, academia, or government.

**Sample Plan of Study**

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Semester/Term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS-GA 1001</td>
<td>Introduction to Data Science</td>
<td>3</td>
</tr>
<tr>
<td>DS-GA 1002</td>
<td>Probability and Statistics for Data Science</td>
<td>3</td>
</tr>
<tr>
<td>Elective</td>
<td>(Data Science, General, Industry, or Track)</td>
<td>3</td>
</tr>
<tr>
<td><strong>Credits</strong></td>
<td></td>
<td><strong>9</strong></td>
</tr>
<tr>
<td>2nd Semester/Term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS-GA 1003</td>
<td>Machine Learning</td>
<td>3</td>
</tr>
<tr>
<td>DS-GA 1004</td>
<td>Big Data</td>
<td>3</td>
</tr>
<tr>
<td>Elective</td>
<td>(Data Science, General, Industry, or Track)</td>
<td>3</td>
</tr>
<tr>
<td><strong>Credits</strong></td>
<td></td>
<td><strong>9</strong></td>
</tr>
<tr>
<td>3rd Semester/Term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS-GA 1006</td>
<td>Capstone Project and Presentation</td>
<td>3</td>
</tr>
<tr>
<td>Elective</td>
<td>(Data Science, General, Industry, or Track)</td>
<td>3</td>
</tr>
<tr>
<td>Elective</td>
<td>(Data Science, General, Industry, or Track)</td>
<td>3</td>
</tr>
<tr>
<td><strong>Credits</strong></td>
<td></td>
<td><strong>9</strong></td>
</tr>
<tr>
<td>4th Semester/Term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective</td>
<td>(Data Science, General, Industry, or Track)</td>
<td>3</td>
</tr>
<tr>
<td>Elective</td>
<td>(Data Science, General, Industry, or Track)</td>
<td>3</td>
</tr>
</tbody>
</table>

**Learning Outcomes**

Upon successful completion of the program, graduates will have:

1. The knowledge and skills needed to develop new methods for data science.
2. The skills needed to use automated analytical tools and cut through the noise to create knowledge from big data.
3. The ability to solve a real-world problem from collecting and processing real-world data, to designing the best method to solve the problem, and finally, to implementing a solution.
4. An awareness of the social and ethical implications of data-driven methods, alongside tools to address and mitigate associated biases.

**Policies**

**NYU Policies**

University-wide policies can be found on the New York University Policy pages (https://bulletins.nyu.edu/nyu/policies/).

**Graduate School of Arts and Science Policies**

Academic Policies for the Graduate School of Arts and Science can be found on the Academic Policies page (https://bulletins.nyu.edu/graduate/arts-science/academic-policies/).