

CS4650 Topic 24:

Data Science Careers

Where I Got This Data:

- [Coursera.org](https://www.coursera.org)
- [Simplilearn.com](https://www.simplilearn.com)
- [Brainstation.io](https://www.brainstation.io)

Introduction

- Data Science is a fairly new field, but it has rapidly grown to be one of the most in-demand technology career path.
- There are several million open jobs in this area, with 20-40% growth rate.
- It is considered the third-best job in the US in 2022.
- The insights gained by Data Science help businesses maximize their profits, investigate new product directions, and develop marketing strategies.
- A Data Scientist is generally a high-level position, so traditionally companies are looking for well-educated individuals.

What Is Data Science?

- Companies have always gathered and analyzed data about their customers.
- Data Science grew out of statistics and data mining:
 - Statistics: the study and manipulation of data, for the purpose of drawing conclusions
 - Data Mining: the process of searching and analyzing raw data to identify patterns and extract useful information
- Data Science includes:
 - Software development
 - Machine learning
 - Research
- Information technology is evolving at a rapid pace
- There is a growing need to make sense of the vast amount of data being generated and collected.

What is a Data Scientist?

- A professional who specializes in analyzing and interpreting data.
- Use their skills to help organizations make better decisions and improve their operations.
- Have a strong background in mathematics, statistics, and computer science.
- Analyze large data sets to find trends or patterns.
- May develop new ways to collect and store data.

Technical Skills

- More than cursory understanding of statistics and statistical analysis
- Mathematics
- Programming
- Engineering
- Data visualization
- Machine Learning
- Deep Learning
- Information Technology
- Familiarity with Hadoop, Spark, Python, R, SQL, Excel, Jupyter Notebook, Tensorflow

Business Skills

- Communication skills
- Storytelling
- Critical Thinking
- Logic
- Curiosity
- Problem Solving
- Teamwork
- Business acumen
- Flexibility

Roles

- Data Scientist
- Data Analyst
- Data Architect
- Data Engineer
- Machine Learning Engineer
- Business Analyst

Data Scientist

- Build models using programming languages such as Python
- Transform these models into applications
- Average salary: \$104K
- Skills: Statistics, mathematics, machine and deep learning, programming, data analysis, big data processes, Hadoop, SQL
- Education: BS, usually MS, professional certificates.
- According to Burtch Works, more than 94% of data scientists have MS or PhD.

Data Analyst

- Use structured data to solve business problems
- Uses SQL, Python, R, statistical analysis and data visualization
- Acquire, clean, reorganize data, search for trends.
- Average salary: \$66K
- Skills: SQL, Python, R, statistics, math, data visualization
- Education: BS in math, computer science, finance or statistics

Data Architect

- Create blueprints for data management systems
- Design plans to integrate and maintain data sources
- Goal is to provide information to employees when they need it.
- Average salary: \$121K
- Skills: Coding languages such as Python and Java, data mining and management, machine learning, SQL, data modeling.
- Education:BS in data or computer science

Data Engineer

- Prepare and manage large amounts of data.
- Develop and optimize data pipelines and infrastructure.
- Build data warehouses.
- Prepare data for data scientists and business analysts to use.
- Average salary: \$97K
- Skills: Java, Python, SQL, NoSQL, Hadoop, database design
- Education: BS in math, science, or business-related field.

Machine Learning Engineer

- Not entry-level position! Can build toward from data scientist or engineer.
- Use algorithms to interpret data and build accuracy.
- Average salary: \$109K
- Skills: Tools such as Spark, Hadoop, R, Kafka, Tensorflow, Google CMLE.
Data structures and modeling, quantitative analysis, computer science.
- Education: MS or PhD

Business Analyst

- Use data to form business insights and make recommendations.
- Identify issues in how a business is organized or run.
- Average salary: \$76K
- Skills: SQL, Excel, data visualization, financial modeling, data and financial analysis, business acumen
- Education: BS in economics, finance, computer science, statistics, business

Path to a Career

- Education: BS or MS in data science, business, economics, statistics, math, information technology.
- Online courses and professional certificates: These can be helpful, and typically take a few months to complete.
- Bootcamps: May take a few weeks to a few months, can be expensive. Typically used when changing a career path.
- Entry-level job or internship: Apply for jobs that specifically cater to those starting out in data science. Some job seekers report applying to hundreds of jobs before receiving an interview, but you are in demand, so keep at it!
- Evidence of requisite skills often outweighs mere credentialism.

Work on Data Science Projects

- Practice using what you have been learning, put them to use.
- Develop a working knowledge of these skills.
- Take on projects that draw on a wide range of skills, don't just do the same type of project over and over. Breadth of experience is important.
- Use different languages, Python, R, Java
- Touch on different stages of the process: initial research, defining, collecting data, cleaning, testing, etc.
- Take on projects that demonstrate a wide range of skills and feature different datatypes.
- Document the stages and what you learned.

Work on Data Science Projects

- Add descriptions to your projects, highlighting your process, what you have learned, the conclusions you have drawn, the experiments you have tried. Feature your creativity.
- Create and apply custom algorithms to analyze and model your data.
- Package your data into easy-to-read visuals
- Practice presenting your results to friends.
- Join Kaggle, then work on several of the projects they offer.

Visualizations and Presentations

- Use several visualization programs, and build some visualizations from scratch.
 - Excel
 - Tools we've used in class (pandas, scikit-learn, seaborn)
 - Tableau
 - Plotly
 - PowerBI
 - Infogram
- Practice presenting your findings to an audience
 - Develop your communication skills

Personal Portfolio

- Build a polished portfolio showcasing the impressive range of your projects.
- Your portfolio may be the most important contribution to your job hunt.
- Display your work on Github as well as a personal website.
- Showcase a wide range of techniques in your projects
- Accompany your data with compelling narrative and context
- Highlight a few key pieces related to your preferred role

The Interview

- Practice your communication skills
- Practice with a non-technical friend
- Describe projects you have built:
 - Talk through your decisions about the tools you chose
 - Explain why you coded the algorithm in a certain way.
- Prove you are familiar with the languages and systems you'll be using on the job.

Conclusion

- The Data Science field has lots of opportunity.
- Receiving a quality education is a great benefit.
- Working on data science projects is a second great benefit.
- Build an impressive portfolio!