Applied Data Science
using Python

Taught by: Bedoor AlShebli

Hi! :)

Who am I?

- **Bedoor AlShebli**
- **Email:** bedoor@nyu.edu
- **Physical Office:** A5-1119
- **Office hours:** email me
- **Lecture times:** Tues and Thurs
  - Sec 1→ 12:45 to 2:00 pm
  - Sec 2→ 2:10 to 3:25 pm

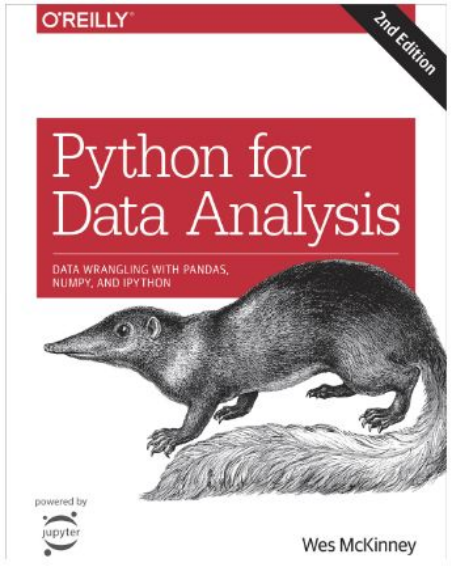
Who's the TA?

- **Sara Saboor**
- **Email:** ss17229@nyu.edu
- **Physical Office:** A5-1172A
- **Office hours:** Tues 3:30-4:30
- **Recitation time:** Wednesdays
  - Sec 1→ 12:45 to 2:00 pm
  - Sec 2→ 2:10 to 3:25 pm

Course Description

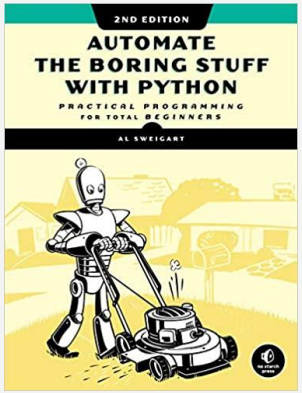
Textbooks:

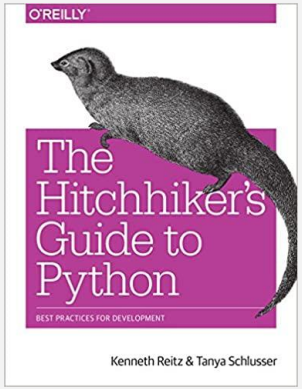
- Python for Data Analysis (required)
- Many free books online as well
- I'll be introducing recommended books in Brightspace



Prerequisites:

- You are expected to have some basic statistics knowledge
- You are expected to know all material covered in Problem Set 0 (ungraded) which will be covered in Recitation 1.





Course Description

What do you expect to get out of this course?

- To be able to apply:
  - Data / Statistical analysis
  - Information visualization
  - Basic machine learning
  - Basic text analysis

… using popular Python toolkits such as pandas, matplotlib, scikit-learn and more to gain insight into any data.

Why Python?

1. It’s easy to learn!

   → Now the language of choice for all top Computer Science programs

2. Full featured

   → Not just a statistical language… it has full capabilities for data acquisition, cleaning, databases, high performance computing, and much much more.

3. Strong Data Science Libraries

   → Example, the SciPy Ecosystem.

Final Grade Breakdown

11 Problem Sets	66%
In Class Participation	5%
Recitation Participation + Submission	5%
Final Project	24%

Attendance is mandatory for both lectures and recitations. Class and Recitation Participations are based on attendance, participation, as well as submission of in class exercises. **Any absence needs to be cleared by the professor prior to class.**

Note: When on zoom, cameras must be turned on please for attendance to count.

A	A-	B+	B	B-	C+	C	C-	D	F
[95,100]	[90,95)	[85,90)	[80,85)	[75,80)	[70,75)	[65,70)	[60,65)	[55,60)	[0,55)

Final Project Grade Breakdown

Throughout the course, you will be submitting and presenting 4 different milestones, completing a real-world data science project. You will be working in groups of 3 (or 4), assigned by us (maybe) based on your background.

Milestone 1: Project Proposal (Hypothesis and Dataset Description) (4%)

Milestone 2: Data Visualizations (5%)

Milestone 3: Preliminary Results Presentation (5%)

Milestone 4: Final Project Report + Presentation (10%)

Let’s get started!

What is Data Science?

Top most in-demand jobs in 2023 : Data Scientist, Business Scientist and Data Engineer

Top most in-demand jobs in 2023 : Data Scientist, Business Scientist and Data Engineer

Mohammed Karimkhan Pathan
AI Team Lead | Sr Data scientist | Data science consultant

24 articles + Follo

February 25, 2023

Open Immersive Reader

Data Scientist

A data scientist symbolizes to organizations a gaping hole: a magic that can tu big data into big gold by making sense of vast amounts and multiplicity of senseless bits and bytes (or zettabits and petabytes?). The data scientist is a savior who (if found) can solve all big data problems, so companies will not ha to worry about figuring out how to do it themselves, all they need is to catch t or three really good data scientists, no matter what they are.

The field of data science has seen tremendous growth over the past few years, and this trend is expected to continue. Companies are increasingly relying on data-driven insights to make informed decisions, and the role of data scientists in providing these insights is critical.

Best Data Science Jobs in 2023

The 365 Team • 2 May 2023 • 5 min read

Deciding on a career path in data science and analytics can be quite challenging. Even if you have a relevant degree, along with additional qualifications from data science courses, the job market trends and industry demands are progressively shifting, and so are the most promising career tracks in the field.

So how do you navigate this ever-changing landscape and choose the best data science job?

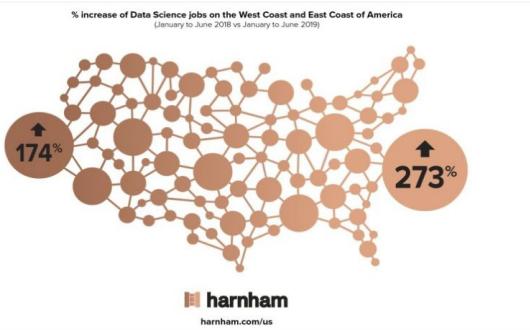
We have looked into data from Glassdoor, Indeed, PayScale, and LinkedIn to pinpoint some essential aspects like salary, job satisfaction rates, job availability, and required skills for some of the top careers in data science today. Based on these findings, we have selected 5 of the hottest jobs in data science that we'll discuss in detail in this article:

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

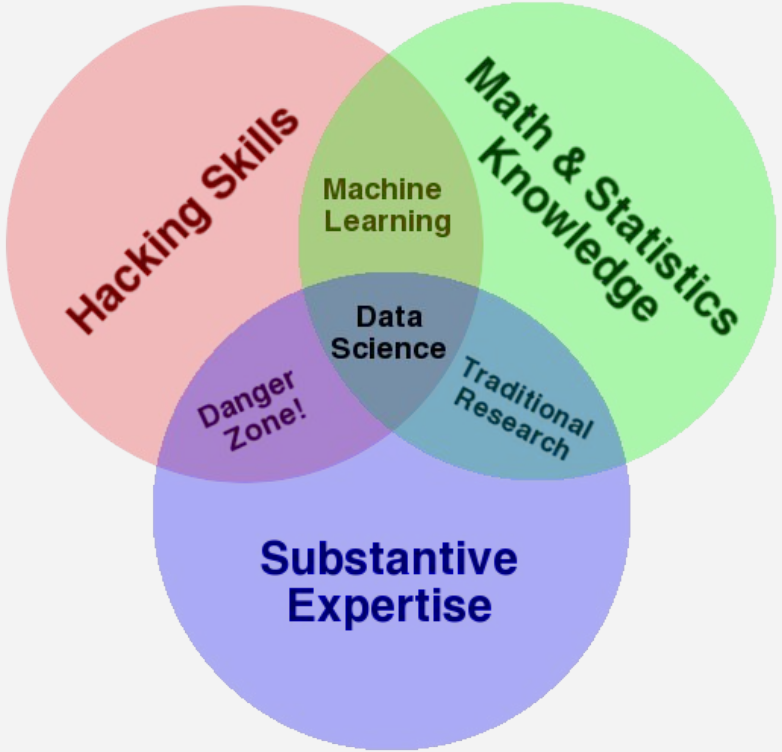
Looking for a New Career? Data Science is Booming

e field is growing rapidly, and not just in Silicon Valley. The number of data science jobs on the East Coast increased 273 percent in 2019, compared to 174 percent on the West Coast.

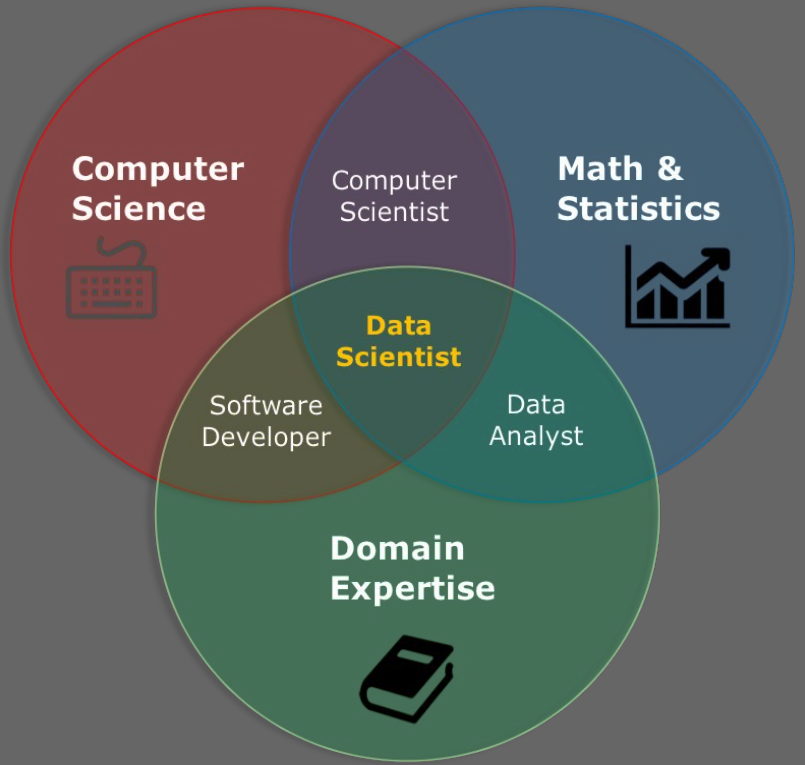
By Angela Moscaritolo Updated December 16, 2019



Drew Conway's Data Science Venn Diagram



Updated Data Science Venn Diagram



50 years of Data Science

David Donoho

Sept. 18, 2015
Version 1.00

Abstract

More than 50 years ago, John Tukey called for a reformation of academic statistics. In 'The Future of Data Analysis', he pointed to the existence of an as-yet unrecognized *science*, whose subject of interest was learning from data, or 'data analysis'. Ten to twenty years ago, John Chambers, Bill Cleveland and Leo Breiman independently once again urged academic statistics to expand its boundaries beyond the classical domain of theoretical statistics; Chambers called for more emphasis on data preparation and presentation rather than statistical modeling; and Breiman called for emphasis on prediction rather than inference. Cleveland even suggested the catchy name "Data Science" for his envisioned field.

David Donoho's "50 Years of Data Science"

An Introduction to the Python Programming Language

But First...

1. Go to Google Drive: drive.google.com
2. Create a folder called “Applied Data Science”
3. Go to the course website, and download today’s Jupyter Notebook, titled “Lecture 1: The Python Programming Language - Functions.ipynb”
4. Upload the file into your “Applied Data Science” folder on Google Drive.
5. Right click on the Jupyter Notebook file after you have uploaded it, and choose to open it with “Google Colaboratory”

Jupyter Notebook and Google Colaboratory

Jupyter Notebook: A very flexible tool commonly used to create *readable analyses*, because one can keep code, images, comments, formula and plots together in one place!

Google Colaboratory:

a. Google Colaboratory are Jupyter Notebooks hosted by Google Research.

b. It allows anybody to write and execute python code via the browser

c. Most libraries are already installed so you won't need to manually do that