

Program Introduction

Welcome to Data Science!



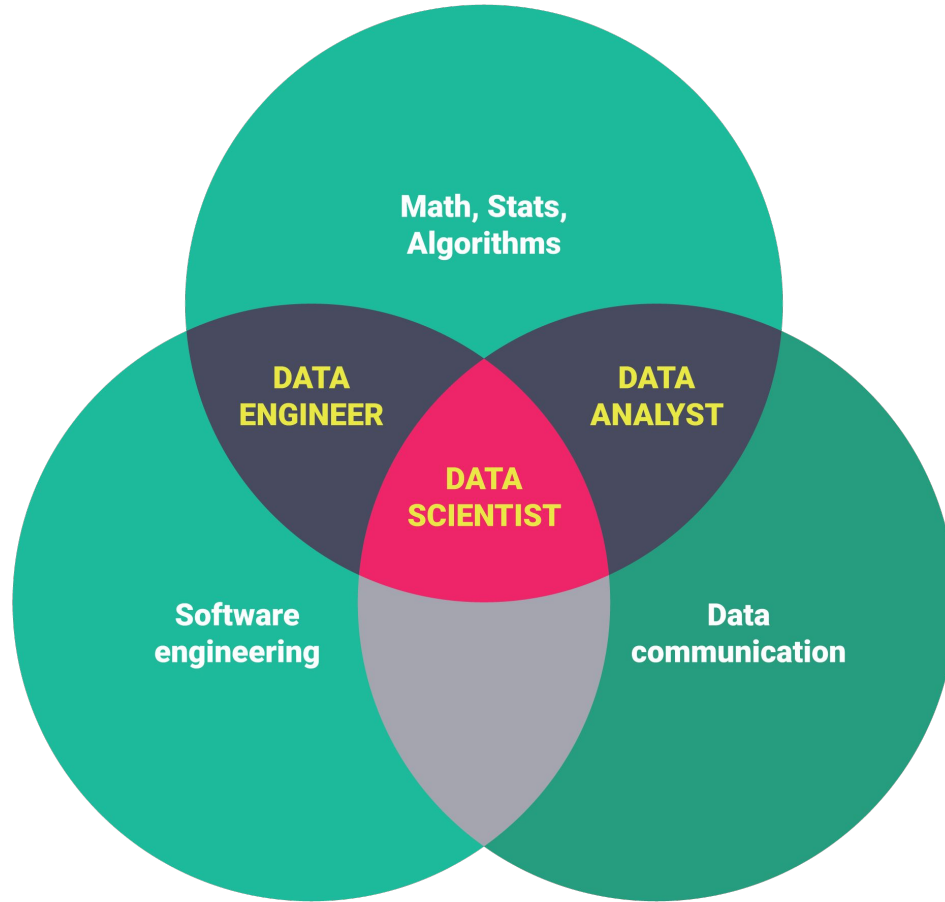
Agenda

1. **What is data science?**
2. **Five modules**
3. **Day-to-day activities**
4. **Start thinking about capstone!**
5. **Keycard setup**

1. What is *data science*?

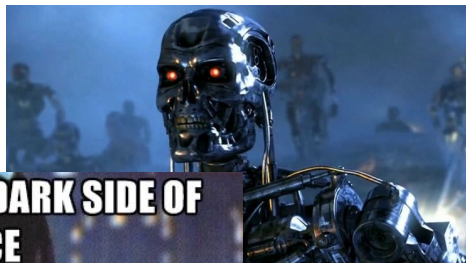
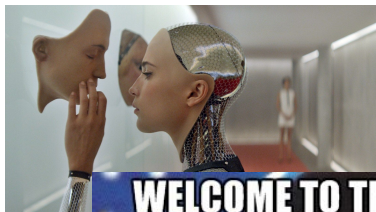
“A data scientist does **model-driven analyses of our data**; analyzes to improve our planning, increase our productivity, and develop our deeper levels of subject matter expertise. A data scientist works at the tactical, operational, and strategic levels, **sharing insights with the business**.”

— Chris Pehura, Practice Director,
Management Consultant at C-SUITE DATA

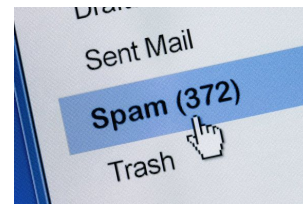
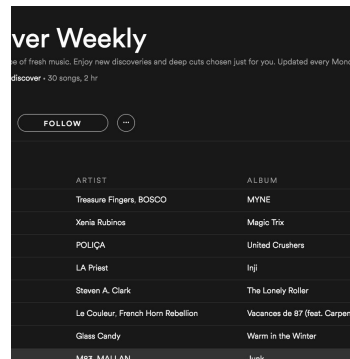


More specifically...

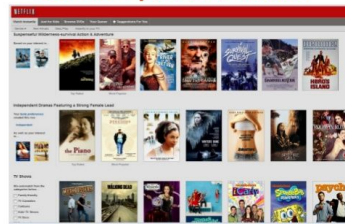
Not this (mostly...)



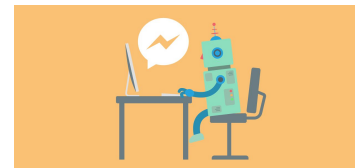
Yes this



Genres - personalization



NETFLIX



2. ***Five modules*** of our data science curriculum

A **module** is one three-week section of the curriculum

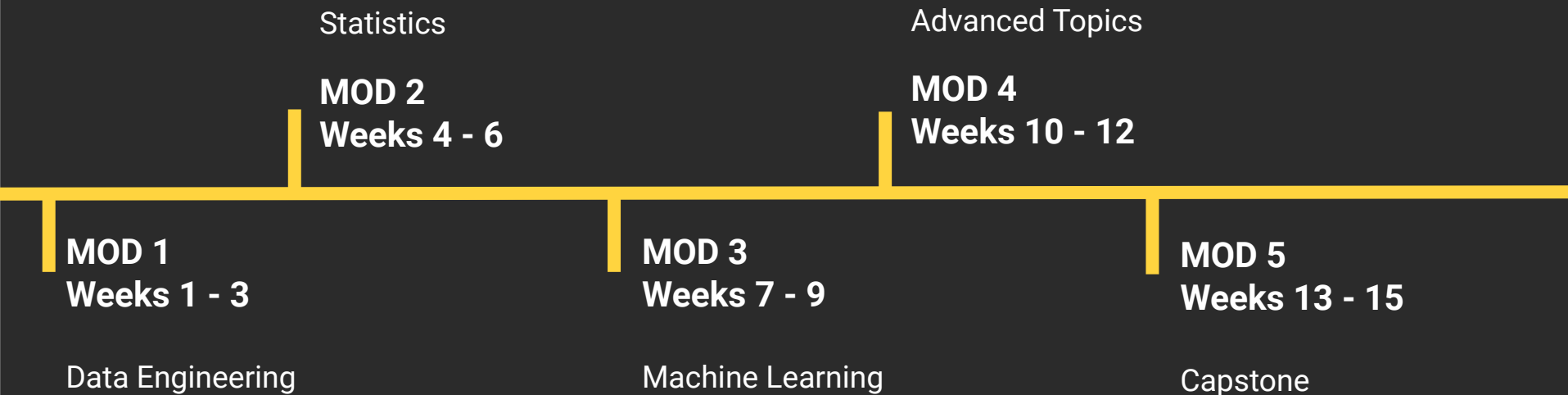
Weeks 1 & 2: * lectures, daily checkpoints, lessons on Learn.co

Week 3: project week

** Mods 1-4 only. Mod 5 is 100% project time!*



Overall Module Timeline



Module 1: Data Engineering

Students Will Be Able To

Read and write data

In CSV, SQL, and JSON formats, using files, databases, APIs, and web scraping

Clean and transform data

Calculate descriptive statistics

Create visualizations

Module 2: Statistics

Students Will Be Able To

***Calculate inferential
statistics***

***Perform hypothesis tests
and A/B tests***

***Run a linear regression
analysis***

To gain insights into past data

Module 3: Machine Learning

Students Will Be Able To

Identify use cases

For regression vs. classification models

Build a supervised machine learning model with scikit-learn

To make predictions about future data

Interpret metrics

To describe the performance of a model

Tune hyperparameters

To optimize model performance

Module 4: Advanced Topics

Students Will Be Able To

Explore advanced data processing techniques

Including principal component analysis, natural language processing, and time series analysis

Identify use cases

For supervised vs. unsupervised learning

Build an unsupervised machine learning model

To categorize data without “ground truth” labels

Use advanced machine learning libraries

For big data (PySpark), deep learning (TensorFlow), recommendation systems, and time series modeling

Module 5: Capstone

Students Will Be Able To

***Complete the full
CRISP-DM process on an
individual data science
project***

- **Business Understanding**
- **Data Understanding**
- **Data Preparation**
- **Modeling**
- **Evaluation**
- **Deployment**

3. Day-to-day ***activities***

Week 1 - 2



Lecture

Classroom format

Content will be sent out
on Slack in advance

Bring your laptop in case
there are code exercises

Lessons on Learn.co

This is our “textbook”

Our lesson order will
differ somewhat

Work will *not* be saved by
default





Pair Programming

“Driver” and “navigator”
work together on one
computer to complete
labs or mini-projects

Daily Checkpoints

End-of-day challenges that sum up what you need to know

Varying lengths and levels of difficulty





Code Challenges

End-of-module challenges that simulate real-world workflows

Evaluating Student Progress

Multiple opportunities to demonstrate your skills

In order to advance through the program, we need to know that you have mastered the essentials

We will look holistically at the daily checkpoints and code challenges (and possibly additional oral or project-based assessments) to determine if you are ready to move forward

This pass/fail assessment happens once per module



Week 3

Group Projects

Groups of 2+ students

Tackling real problems

Developing a portfolio of projects



4. Start thinking about ***capstone***

Immerse yourself in the data science community

Learn about new tools, techniques and libraries being used to solve current problems in data science!

Blogging

The Gradient
thegradient.pub

Locally Optimistic
locallyoptimistic.com

Uber Engineering AI
uber.com/category/articles/ai/

Podcasts

Not So Standard Deviations

Linear Digressions

DataSkeptic

Twitter

Jake VanderPlas
Author of Python for Data Science
[@jakevdp](https://twitter.com/jakevdp)

DJ patil
Former U.S. Chief Data Scientist
[@dpatil](https://twitter.com/dpatil)

Rachel Thomas
Director of USF Center for Applied Data Ethics
[@math_rachel](https://twitter.com/math_rachel)

5. Keycard setup

1. Download the WeWork App
2. LOGIN → FORGOT PASSWORD
3. Reset password using your email address
4. Login using your new password
5. Follow instructions to register keycard
6. If you get an error message, send us a Slack message and we'll take you to the WeWork community floor



6. *Introductions!*

1. Introduce yourself to the person sitting next to you.
2. Find out their...
 - a. Name (or preferred name).
 - b. What they did before this.
 - c. Why they are interested in data science!
3. Be prepared to **introduce your partner to the class!**
4. Take the next five minutes to do this!