# Installation
*Get the required software*

**1.**



**Anaconda**



https://www.anaconda.com/distribution/
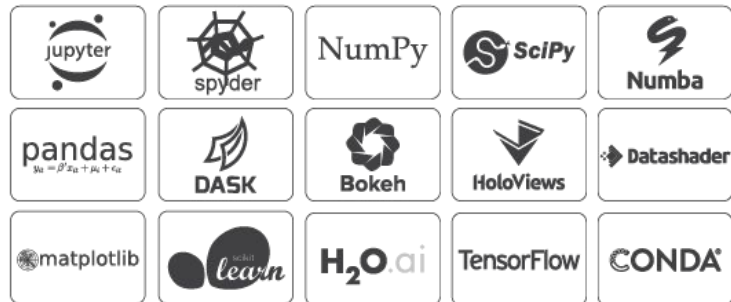
**2.** Open **Terminal (**MacOS**) / Command Prompt (**Windows**),
Type and enter:** jupyter notebook

# INFO 1998: Introduction to Machine Learning

**CDS Education**

We explore, learn, and educate big minds.

# Lecture 1: Introduction

**INFO 1998: Introduction to Machine Learning**

**CDS Education**
We explore, learn, and educate big minds.

# Agenda

1. **Meet the Team**: Who are we?
2. **Introduction**: What is Data Science / Machine Learning?
3. **Course Syllabus**: What will we learn?
4. **Course Logistics**: How will we learn?
5. **Getting Started**: Software & Demo

# Who are we?

**Cornell Data Science**

| Project Team | Community Outreach |
|---|---|
| Intelligent Systems | Education |
| Insights | INFO 1998 |
| Data Engineering | Tech Talks |
| Algorithmic Trading | Online Tutorials |
| Education | |

# Course Manager
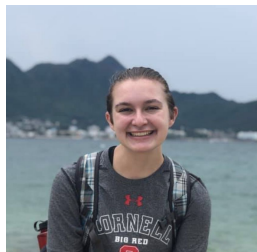
*Who you'll have to bear with*


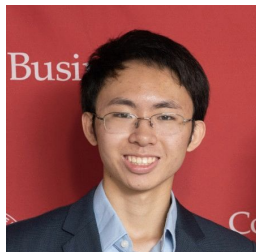
**Emily Weed**

Stats '22

emw232@cornell.edu

*Took INFO 1998 in Spring 2020*
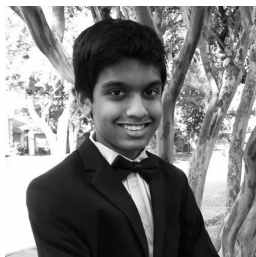
# Course Instructors

*Backbone of INFO 1998*
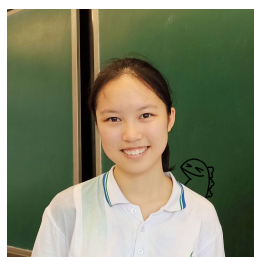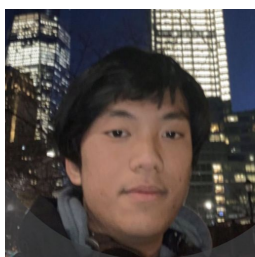
**Sam Cobado**
ECE '22

**Jerry Sun**
CS '23

**Kevin Jiang**
ORIE '23

**Varun Gande**
CS & ORIE '24

**Vivian Chen**
CS & Math '25

**Vincent Fong**
CS & ORIE '24
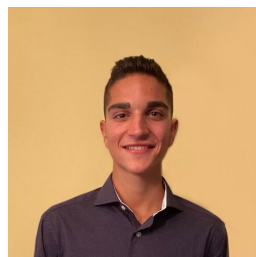
**Jacob Mayourian**
ORIE '25

**Kelly Lu**
Math '23
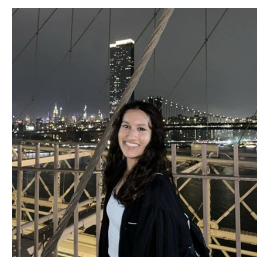
**Neha Kulshreshtha**
ORIE '25

# What is Data Science?

# What is Data Science?

"By "Data Science", we mean almost everything that has something to do with data: Collecting, analyzing, modeling...... yet the most important part is its applications --- all sorts of applications."

*Journal of Data Science*

# What is Data Science?

*Conway's Data Science Venn Diagram*
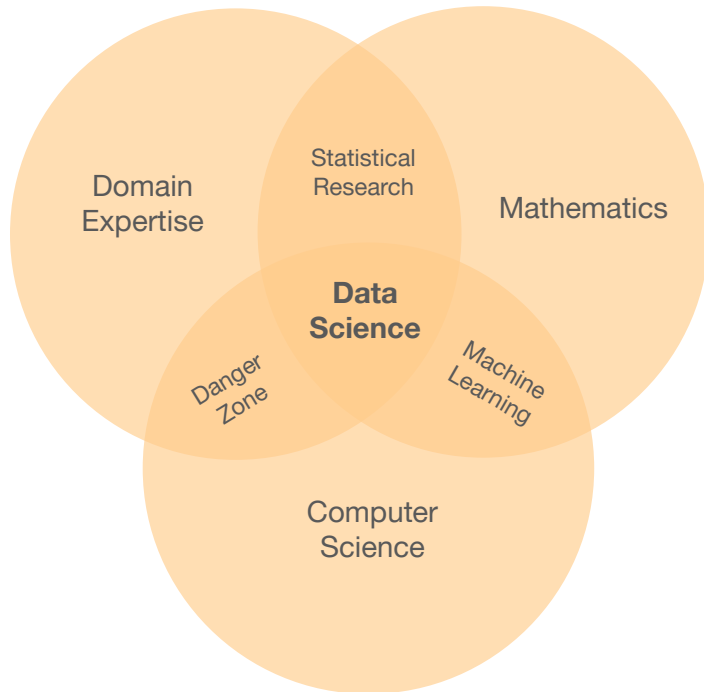
# What is Data Science?

*Conway's Data Science Venn Diagram*
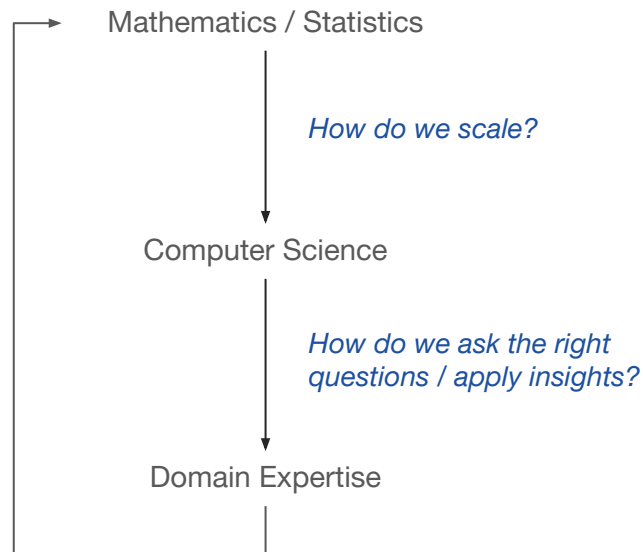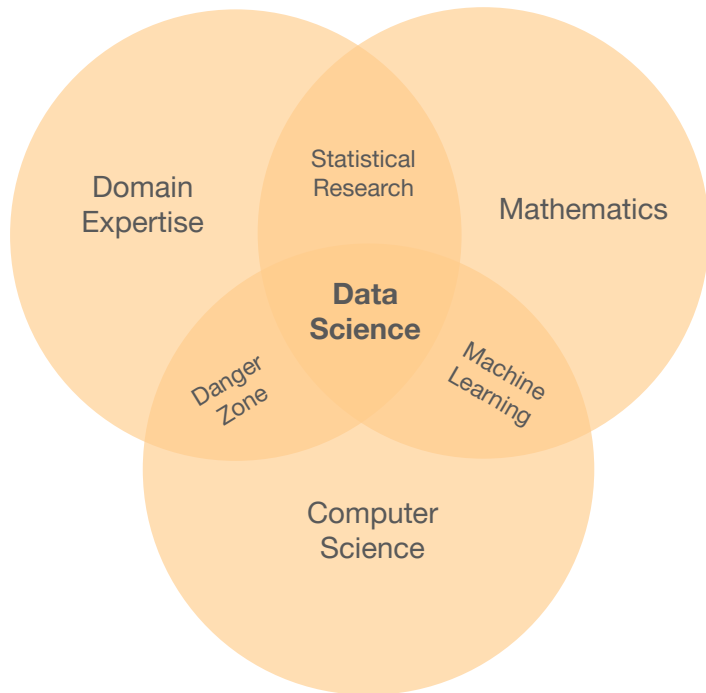
Data Science ≠ Machine Learning

# Applications of Data Science

*We'll be back to this slide!*

| Predictive | Stock Prices |
| --- | --- |
| | Netflix Recommendations |

| Preventive | Medical Diagnosis |
| --- | --- |
| | Social Impact Analytics |

| Real-Time | Digital Advertising |
| --- | --- |
| | Autonomous Vehicles |

# Getting to know your classmates
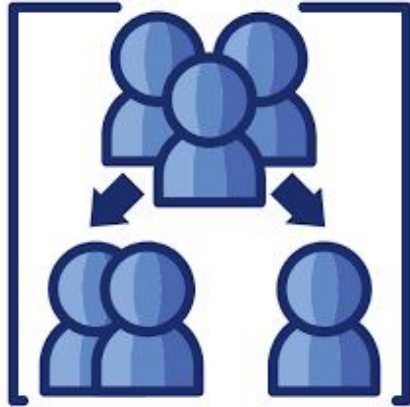
**Spend 5 minutes going over the following:**

- Name
- Major
- Why you are taking this course
- An application of data science you find interesting

# What Do You Get Out of This?

*What you will have accomplished by the end of this?*

# Course Objectives and Syllabus

*What you should aim to understand by the end of the course*

| OBJECTIVES | SYLLABUS |
|---|---|
| Manipulating Data | Data Manipulation / Visualization *Lectures 1-3* |
| Communicating Data | |
| Understanding of ML as a concept | Fundamentals of Machine Learning *Lectures 4-5* |
| Intuitive understanding of ML models | |
| Implementation of ML models | Supervised Learning *Lectures 6-8* |
| Comfort Using Python | |
| Applications in Industry | Unsupervised Learning *Lecture 9* |
| Project Experience | |

**Syllabus is posted on our website under "syllabus"**

# FAQs

*Is this class a good fit for you?*

## 1) Will I become a Data Scientist / Machine Learning Engineer?

*No, you will not. The course covers a breadth of concepts, helps build intuitive understanding of some models, but does not dive too deep into the mathematical complexities (since this is a 1000-level course). However, feel free to come to office hours if you're interested in learning more.*

## 2) How much time commitment is this course?

*Completely up to you. It's not hard to pass the class if all you want is basic street-fighting machine learning skills, which is fine too – that'll require less than 1 hour per week. If you want to put some more time in and come up with a creative and cool data science project that you can be proud of, that will take more time but is very rewarding!*

## 3) I have no background in CS / Stats – am I underprepared?

*Not at all! We'll teach you everything you need to know, but you may have to spend a little more time getting comfortable with Python. A number of non-STEM graduate students have taken this class in the past to understand basics that they could apply to their research. A large number of freshmen also take the course because they're excited to learn more about the field. TL;DR: If you're interested, give it a shot!*

# Course Logistics

*How is the class structured (and graded)?*

**10 assignments (1 assignment per lecture)**          **60%**

*Drop lowest score!*

**Project**          **40%**

*2-3 students*

**Passing Grade: 70%**

# Project Details

*More details to come on the final project rubric and our website*

## Pre-Processing and Manipulation

*Any necessary cleaning and manipulation of the dataset*

## Visualizations

*At least two visualizations. Visualizations are clearly visible, clean, well-labeled, and serve a clear purpose for your question(s).*

## Models

*Machine learning models that are chosen wisely, implemented correctly, and give meaningful results. For example, you won't get points if you run a linear regression for a classification problem. If applicable, the results of the models are compared.*

**Feel free to stay after class to form groups!**

# Sample Final Projects UPDATE

*"0 – 100, Real Quick" - Drake*

## (1) Predicting Heart Failure

*Fernando Celaya, Ming DeMers, Marcus Posey*

## (2) Predicting the Price of Used Cars

*Grant Rineheimer, Benjamin Tang, Dylan Tom*

# Enrollment

*Let's get this credit*



**Fill out to get a pin**
**https://forms.gle/DCJtUMrkKAADuan47**

# Enrollment

*Let's get this credit*

# 1



**Fill out**
https://forms.gle/DCJtUMrkKAADuan47

**Enroll in Ed Discussion**
**https://edstem.org/us/join/uACyDS**

**You will be added to CMS over the weekend**

# Enrollment

*Let's get this credit*

**1**

**2**

**3**

**Fill out**
https://forms.gle/DCJtUMrkKAADuan47

**Enroll in Ed Discussion**
https://edstem.org/us/join/uACyDS

**You will be added to CMS over the weekend**

**Enroll through Student Center after obtaining pins**

# Enrollment on Student Center

*Let's get this credit*

- **Get enrollment pin via email (sometime this weekend)**
- **Add INFO 1998 Section 602 under Rene Kizilcec**
- **Deadline to enroll is February 25th**

1. Select classes to add - Enrollment Preferences

Fall 2020 | Undergraduate | Cornell University
INFO 1998 - Freshmen Team Projects

**Class Preferences**

INFO 1998-602          Ind Study          🟢 Open

    **Topic**   Intro to Machine Learnig
    **Session**  Project Session Full
    **Career**   Undergraduate

Enrollment Information

- Instructor Consent Required to enroll in this class

**Wait List**  ☐ Wait list if class is full

**Permission Nbr** [    ]    ← pin

**Grading** [ Sat - Unsat (SUV) ▾ ]

**Units** [ 1.00 ▾ ]    ← 1 credit pass/fail

[ Cancel ]  [ Next ]

| Section | Component | Days & Times | Room | Instructor | Start/End Date |
|---------|-----------|--------------|------|------------|----------------|
| 602 | Ind Study | | To Be Assigned | Haym B Hirsh, Karthik Sridharan, David Mimno | 09/02/2020 - 12/16/2020 |

# Where can I find course information?

**Asking Questions:** Ed Discussion

*https://edstem.org/us/join/uACyDS*

**Assignment File & Submitting Assignments:** CMS

*https://cmsx.cs.cornell.edu/*

**Lecture Recordings & Assignment Files:** Course Website

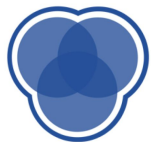*https://cornelldatascience.github.io/info1998/*

# Python Demo

*Feel free to follow along in Jupyter Notebook*

# Next Steps

- **Installation**: Seek help at Office Hours!
- **Assignment 1**: Due at 4:30pm on Wednesday, Feb 23, 2022 on CMS
- **R/Python Workshop:** Fill out poll on ED Stem to pick a day/time
- **Enroll on Student Center:** Will receive a pin through your Cornell email this weekend
- **Next Lecture**: Data Manipulation

**CDS Education**
We explore, learn, and educate big minds.