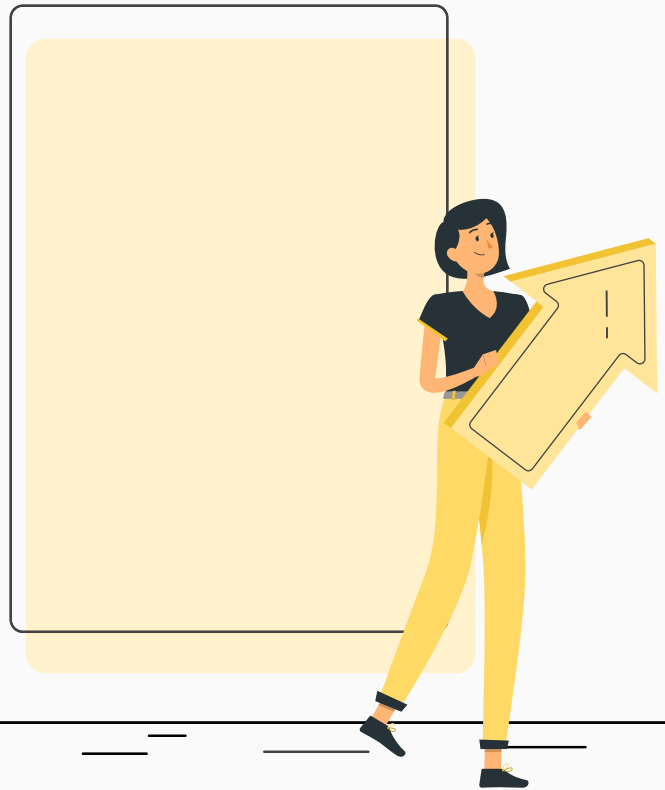
APPLIED DATA SCIENCE II

Course Introduction

Kyle Scot Shank, '14
FA-21



01 INTRODUCTIONS

Let's all say hello to one another!

02 SYLLABUS REVIEW

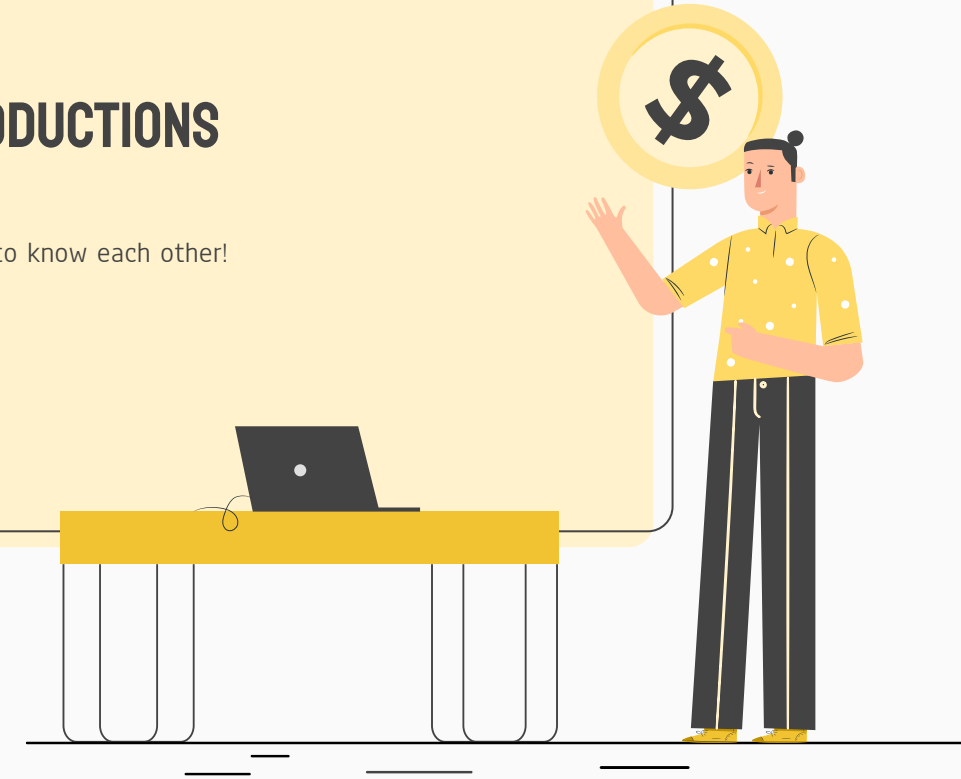
Let's see what's in store for the next ten weeks/

03 Q&A

Let's talk about things!

INTRODUCTIONS

Let's all get to know each other!

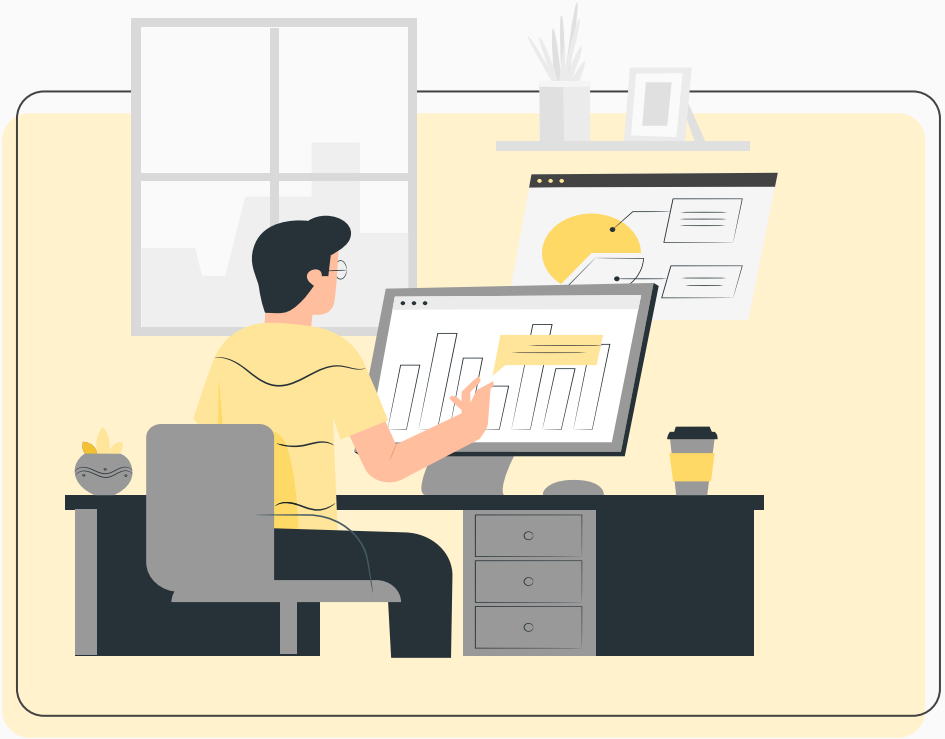


ABOUT KYLE...

- Visiting Faculty Member (this is my second term teaching @ COA!)

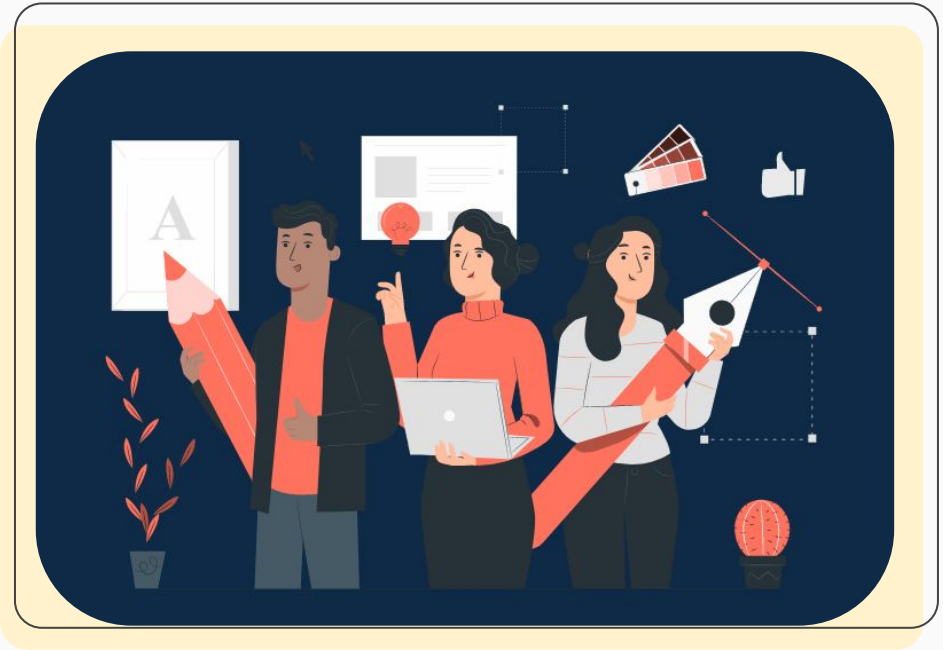
- Currently a quantitative researcher @ FACEBOOK, previously worked in Data Science-y jobs @ The Hershey Corporation, MDI Biological Laboratory, and energy efficiency consulting.

...also bartended in town!



THINGS I RESEARCH...

- How do individuals respond to various different forms of advertising and how can we make those ads more effective?
- How can we quantify and validate metrics related to creativity?
- How can we make housing more equitable and affordable on MDI?
- How can we use offline data from individual, small municipalities to identify broad, state-wide trends?

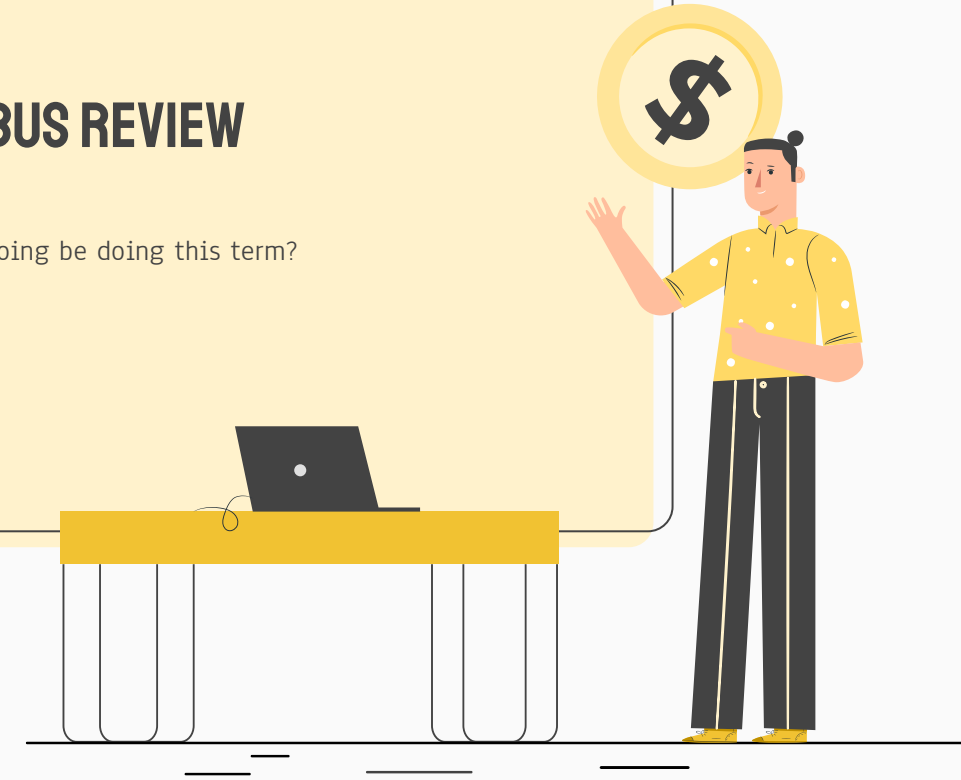


ABOUT YOU!

- What is your name?
- What year are you @ COA?
- What is something you did over the winter break that was awesome?
- What do you want to learn in this course?

SYLLABUS REVIEW

What are we going be doing this term?



COURSE OBJECTIVES

01
UNDERSTAND MODELING

- We are going to really dive into the *why* and *when* of building predictive models. We'll explore what "predictive" means, what "modeling" means, and poke around the philosophical edges a bit.
- We'll also explore the various considerations that one needs to take into account when thinking about generating "predictions" for things - such as confidence, bias, error rates, etc.

02
BUILD A TON OF MODELS!

- We're going to build all different sorts of predictive models and try them out on different data sets and for different situations.

PROGRAM OBJECTIVES

You are here!

APPLIED DATA SCIENCE I

Main Focuses: Data Manipulation, Aggregation, and Visualization

APPLIED DATA SCIENCE II

Main Focuses: Data Modeling, ~~Inference~~, and Prediction

APPLIED DATA SCIENCE III

Main Focuses: Collaboration, Communication

ASSIGNMENTS AND POLICIES

CLASS PARTICIPATION

We want to focus on *presence*, *attention*, and *preparation* - not necessarily just actively talking during class.

10% of total course grade

HOMEWORK ASSIGNMENTS

There will be weekly homework assignments. They will be graded for *completeness*, *thoroughness*, and *correctness*

60% of total course grade

FINAL PROJECT

An in-depth modeling exercise of a data set of your choice - either done in a group or individually.

30% of total course grade

ASSIGNMENTS AND POLICIES (CONT'D)

PLAGIARISM

Don't copy other people's work or use other peoples' work without attribution.

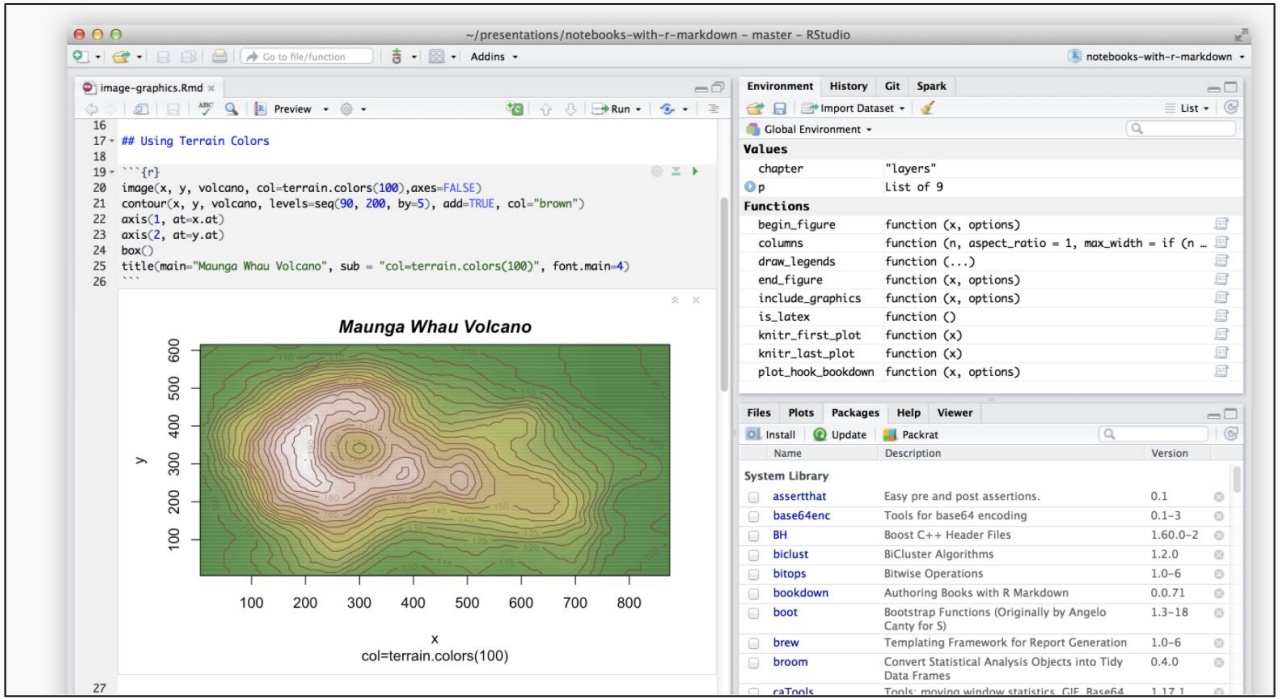
THE INTERNET IS YOUR FRIEND

Modern Software Engineering + Data Science *explicitly* uses open source methods to get work done - so don't feel bad about using Google / StackOverflow (but cite your sources!)

IF YOU'RE FEELING LOST OR STRESSED - REACH OUT

This class might contain a *lot* of new skills being learned all at once - it's okay to feel stressed! Please reach out early with any difficulties and I'll help.

A QUICK NOTE ON HOMEWORK ASSIGNMENTS + RMARKDOWN



RMARKDOWN IS A VERY CONVENIENT WAY TO ITERATIVELY WORK ON YOUR CODING ASSIGNMENTS + PRODUCE PUBLICATION-LEVEL OUTPUTS.

YOUR HW ASSIGNMENTS WILL NEED TO BE SUBMITTED VIA .PDF OUTPUTS FROM RMARKDOWN FILES.

A TEMPLATE HAS BEEN PROVIDED FOR YOU AND IS ON THE GOOGLE DRIVE.

ASSIGNMENTS AND POLICIES (CONT'D)

LATE SUBMISSION POLICY

Check the syllabus, but TLDR:

You are entitled to submitting up to **one** assignment up to **7 days** late without penalty - after that, it's a 5 point (~5%) penalty per day late.

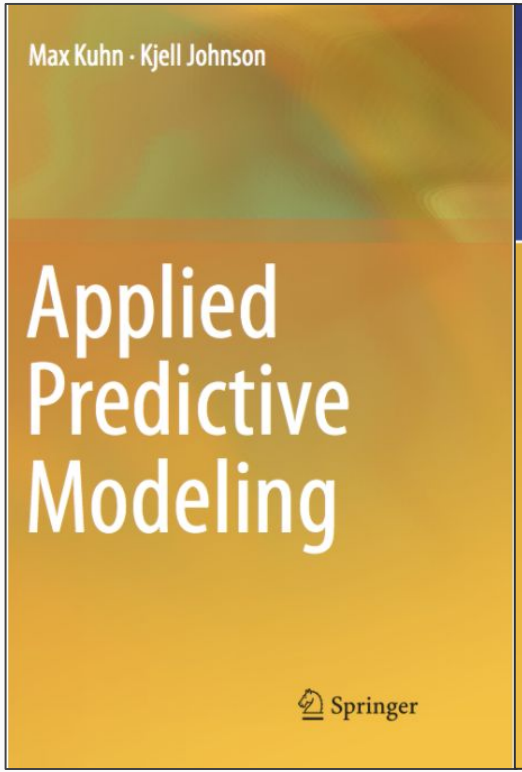
COVID-19 REMOTE INSTRUCTION

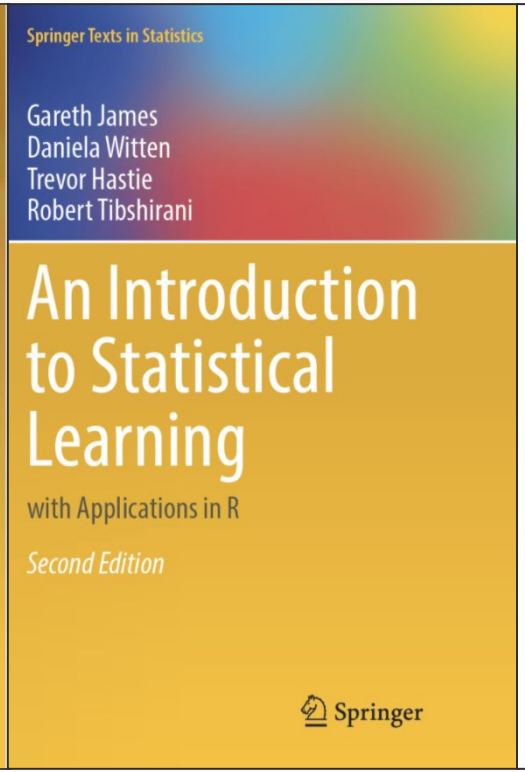
We're going to really try our best to be in-person this term - but COVID-19 is still a thing, If we'll be remote, I'll try to let you know **at least** 24 hours ahead of time and provide a Zoom link.

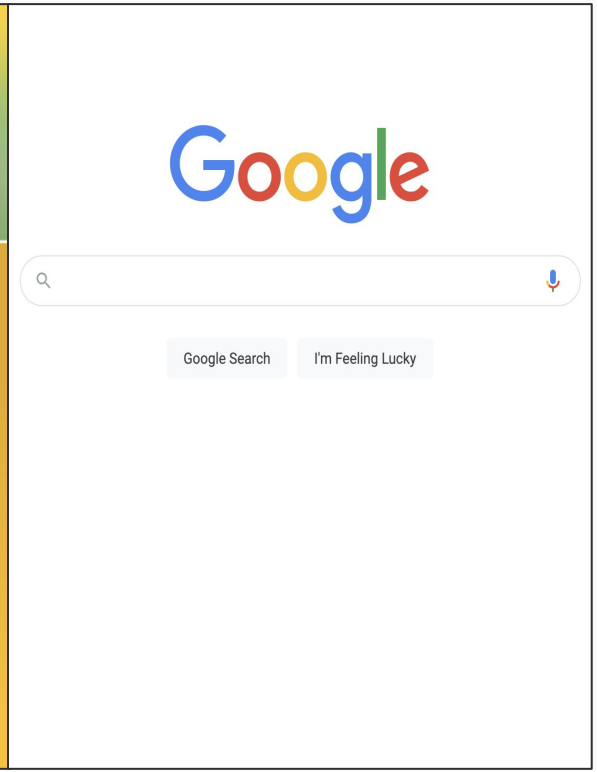
CLASS COMMUNICATION

We're going to lean a bit heavier on Google Classroom for ADS-II: so make sure to check there for announcements each week.

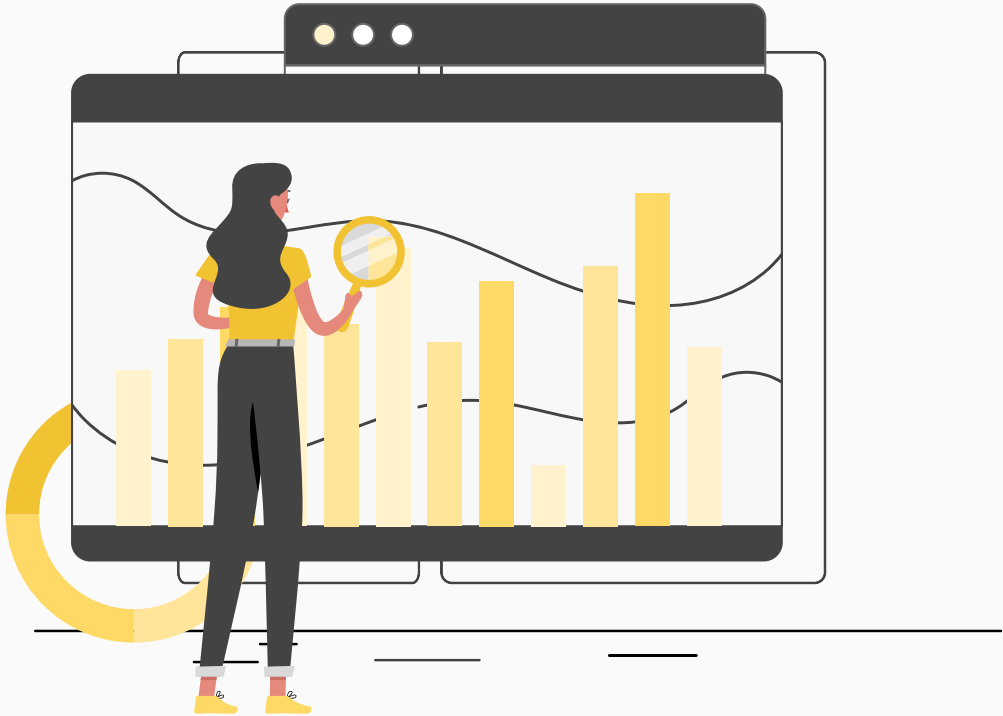
TEXTS + RESOURCES



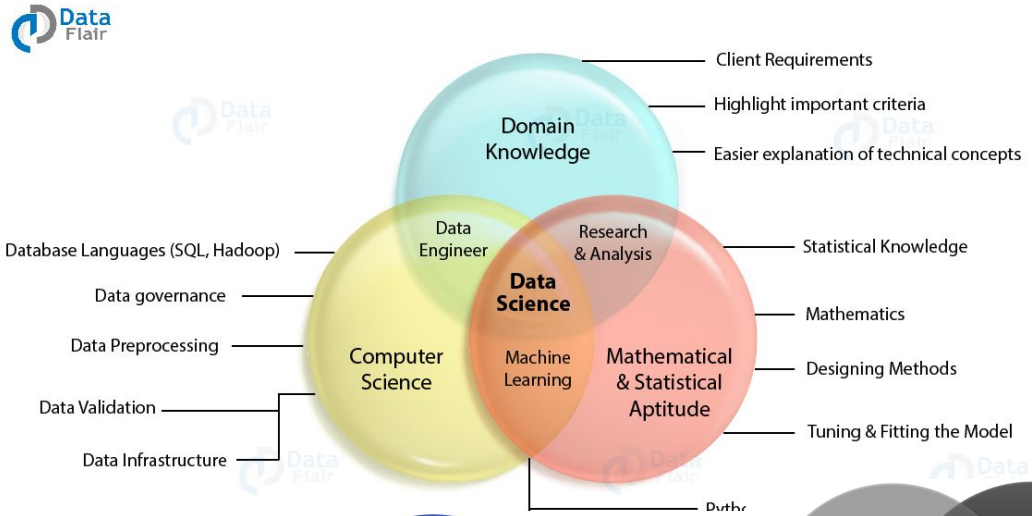


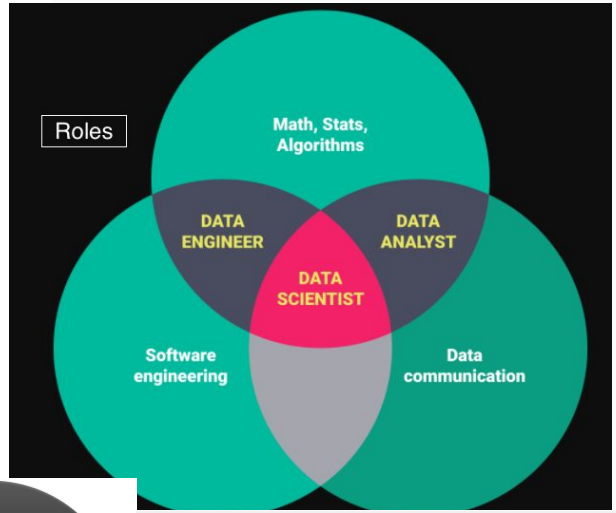


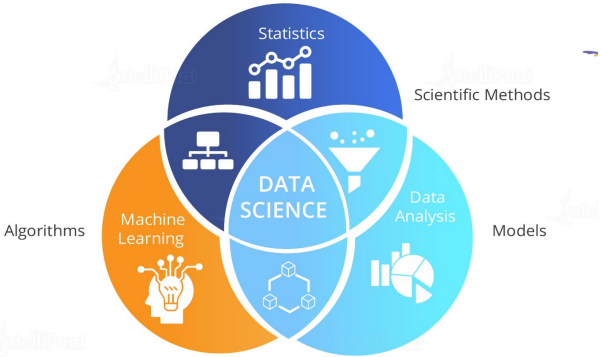
WHAT IS DATA SCIENCE?

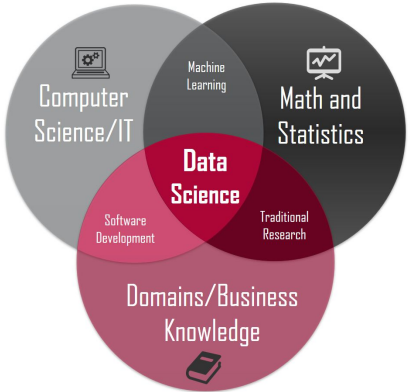


WHAT IS DATA SCIENCE?







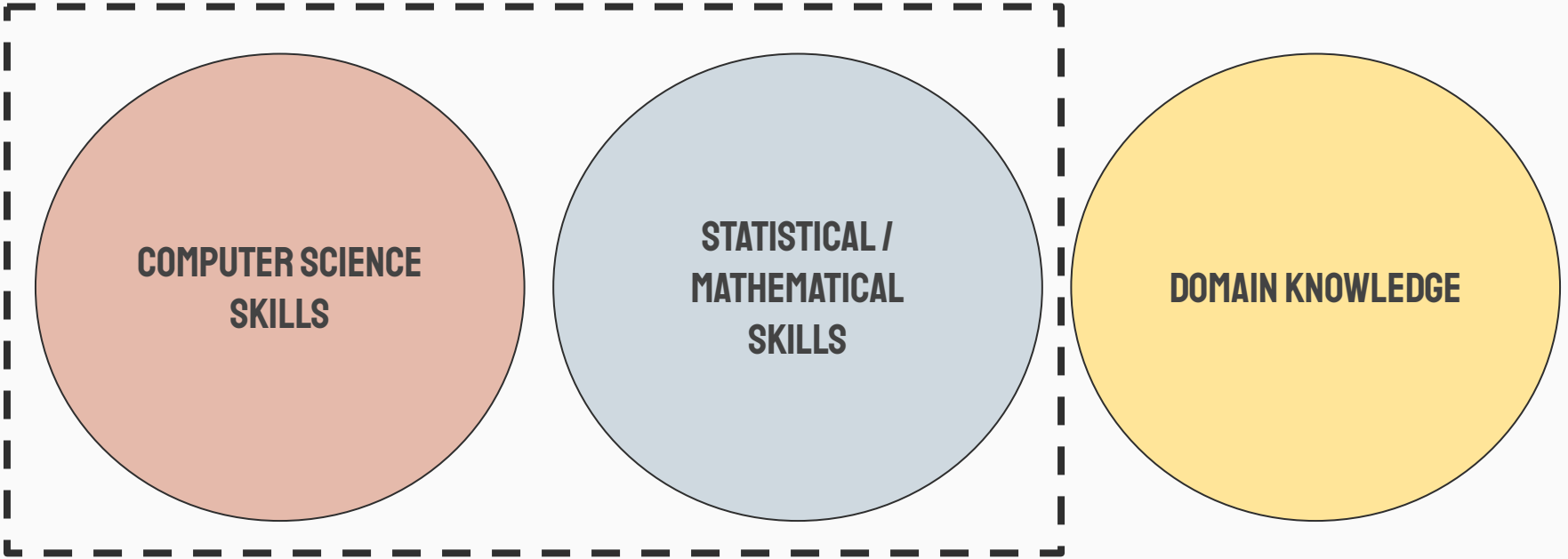


What is Data Science?

Data Science & its scope in IT industry

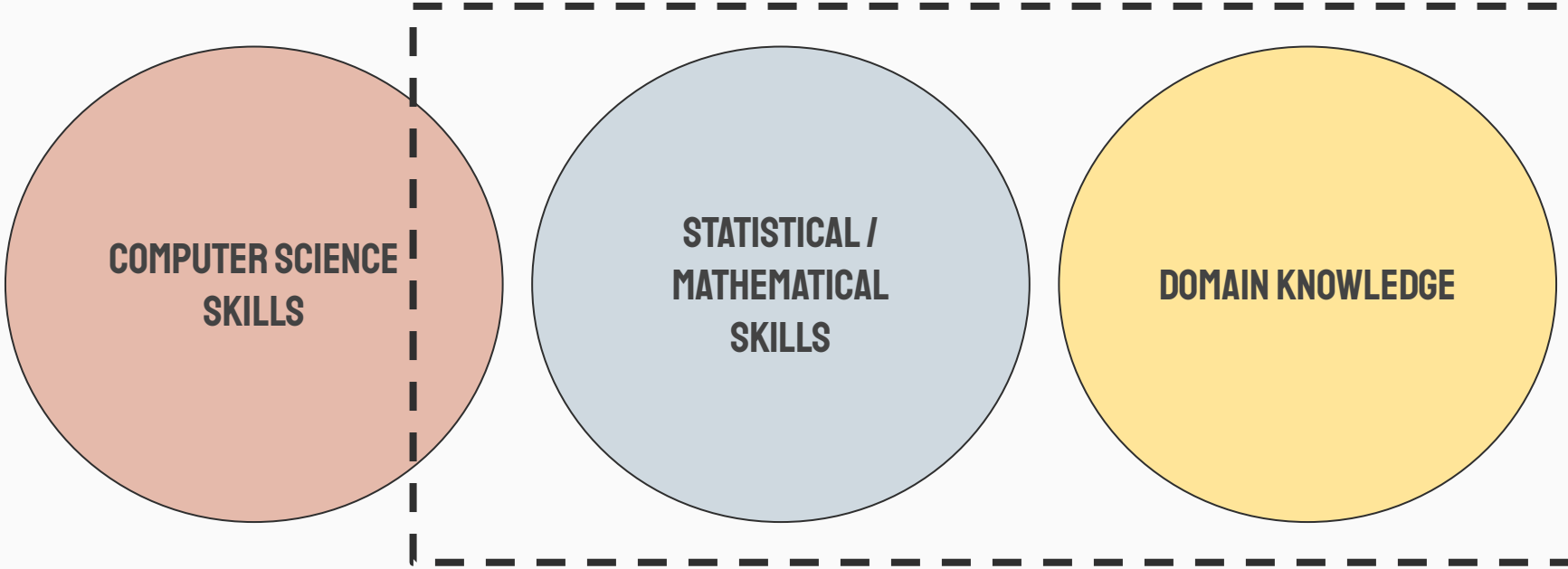
What is the difference between Big Data and Data Science?

WHAT IS DATA SCIENCE?

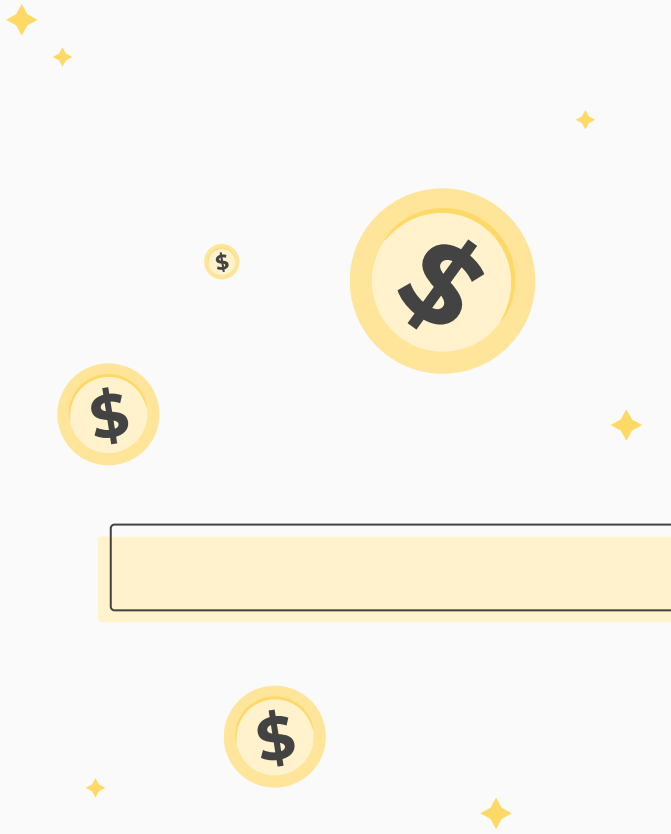


YOU CAN REALLY DO A LOT WITH JUST THIS PART!

WHAT IS DATA SCIENCE?



BUT WHEN IT COMES TO PREDICTIVE MODELING - YOU'LL *REALLY* START TO NEED THAT DOMAIN KNOWLEDGE



"I like to think of data as the new soil;
get in and get your hands dirty."

—DAVID MCCANDLESS

Q&A