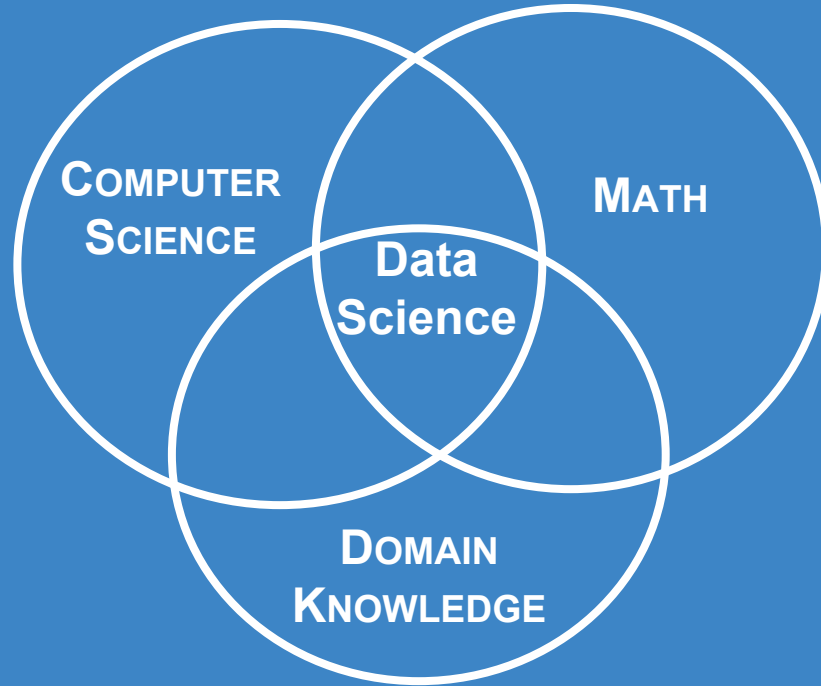# Introduction to
Data Science

Outline

- What is Data Science?
- Why Data Science?
- The Data Science Process
- Course Overview

What is
Data Science?

What is **Data Science?**

Leveraging Computer Science, Math (Mostly stats) and domain knowledge to generate insights and value from data.

COMPUTER SCIENCE

MATH

Data Science

DOMAIN KNOWLEDGE

Why
Data Science?

Why **Data Science?**

Data is growing in terms of size but also in terms of reach (across industries and also our personal lives).

Why **Data Science?**
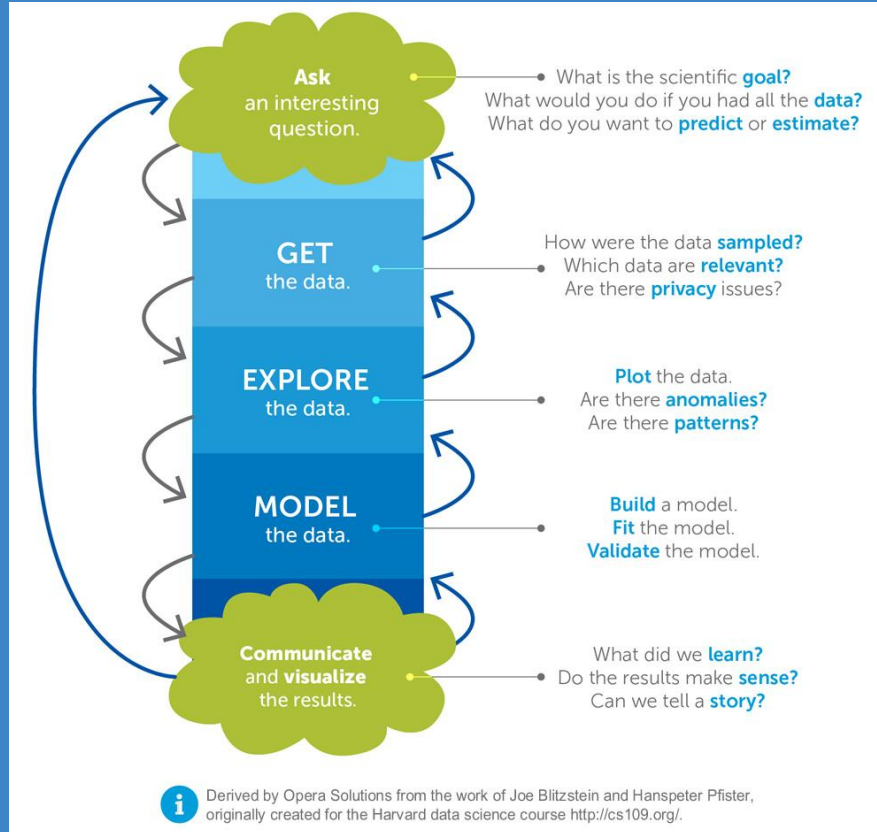
- Impact and Value

- Insights

- Fun

The
Data Science
Process

The **Data Science** Process

i) Define the Problem/Question

ii) Acquire the data

iii) Explore and Prepare Data

iv) Build and Evaluate Models

v) Communicate Results

The **Data Science** Process

Ask
an interesting
question.

What is the scientific **goal?**
What would you do if you had all the **data?**
What do you want to **predict** or **estimate?**

GET
the data.

How were the data **sampled?**
Which data are **relevant?**
Are there **privacy** issues?

EXPLORE
the data.

Plot the data.
Are there **anomalies?**
Are there **patterns?**

MODEL
the data.

Build a model.
Fit the model.
Validate the model.

Communicate
and **visualize**
the results.

What did we **learn?**
Do the results make **sense?**
Can we tell a **story?**

Derived by Opera Solutions from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course http://cs109.org/.

Course Overview

Course Overview (Philosophy)

- Focus on *doing* over everything else
- Adapt and iterate
- Learn Together

- Content:
  - Wide enough to get a sense of the big picture, deep enough to actually follow the process end-to-end.
  - Lean towards the *Practical* side of things

Course Outline

Week 1. Introduction to Data Science and Python

Week 2. Data Exploration and Analysis I (Numpy, Pandas, Matplotlib

Week 3. Data Exploration and Analysis II (Probability & Statistics)

Week 4. Machine Learning I

Week 5. Machine Learning II

Week 6. Applications: NLP

Questions?