Introduction to Data Science

BRIAN D'ALESSANDRO

ADJUNCT PROFESSOR, NYU

FALL 2019

Fine Print: these slides are, and always will be a work in progress. The material presented herein is original, inspired, or borrowed from others' work. Where possible, attribution and acknowledgement will be made to content's original source. Do not distribute, except for as needed as a pedagogical tool in the subject of Data Science.

COURSE FUNDAMENTALS

INSTRUCTOR

Brian d'Alessandro

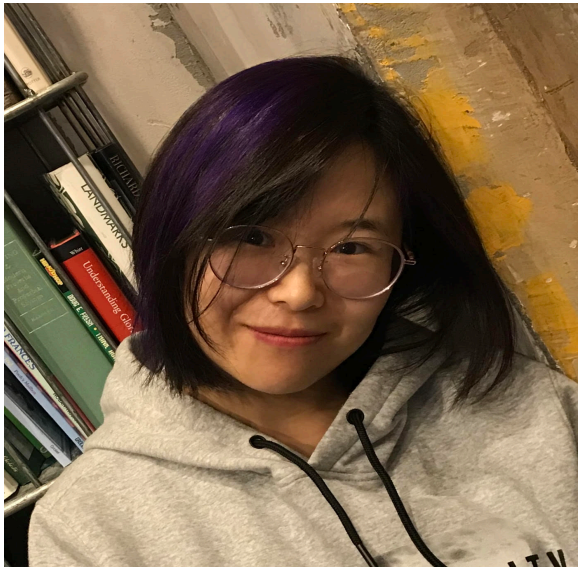
Bio

Education:
Undergrad: Rutgers, Math
Grad: NYU Stern, Statistics

Professional Experience
Capital One
ZocDoc
Facebook
Dstillery (AdTech)
Meetup.com (Social Web)
American Express (Credit/Risk)

Affiliations/Publications
ACM KDD
Big Data Journal
Machine Learning Journal
SIAM

SECTION LEADERS



Nan Wu
Email: nan.wu@nyu.edu

Office hours:
See syllabus

Lee Tanenbaum
Email: leedtan@gmail.com

Office hours:
See syllabus

GOALS OF THIS COURSE

- Understand what a Data Scientist is
- Approach applicable problems data-analytically
- Have hands-on experience mining data

PROGRAMMING!

In order to succeed and participate in this class, You will…

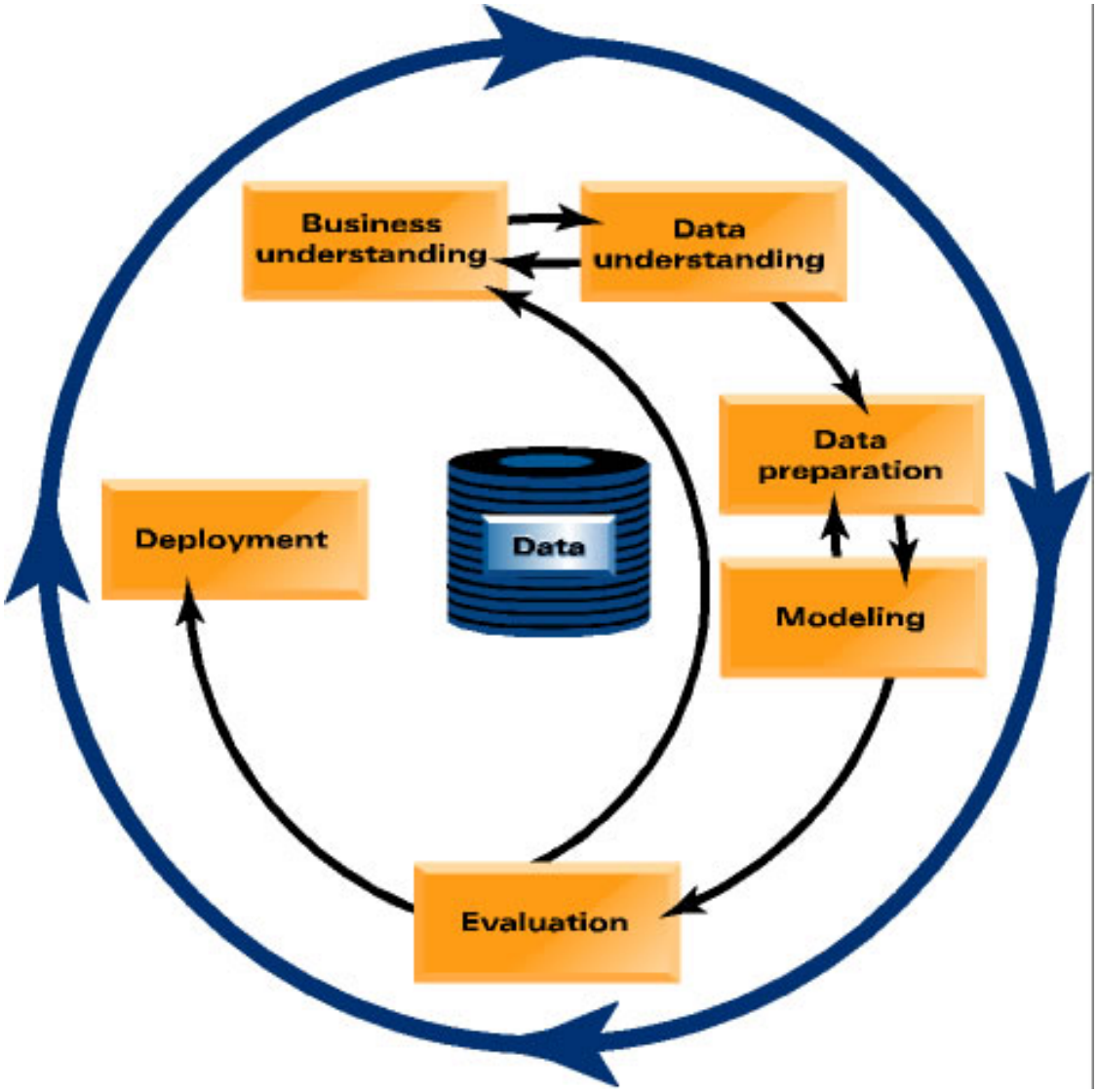
- Need access to a computer with admin privileges
- Have to learn and use the Python programming language.

Please see me after class if this is an issue.

LECTURE OUTLINE

This course will work in the same flow as a typical data mining project.

We'll also peel the layers of data mining like an onion, so the flow might not always be linear.



HOMEWORK

- Will be announced the week they're assigned
  - We'll spend time in class discussing what is expected
- Are all expected to be performed solo
  - Discussions are welcome and encouraged on the course forum
- Will involve Python analysis with discussion of results
  - Open questions will be assigned
  - Code and open answers to be turned in
- Syllabus is the de-facto guide to timing
  - See Syllabus for late policy

QUIZZES AND DISCUSSIONS

- Test your accumulated knowledge throughout the course
- Flex your case scoping muscles
- Keep you engaged and on-time!

FINAL PROJECT

The final project will pull together all of the elements you learn from this class and will simulate the experience of being a professional data scientist. Ultimately, we want you to be able to identify a problem, implement a solution, and demonstrate the value of your solution.

Milestones (due dates will be assigned)

1. Choose a team
2. Pick a dataset and a business problem, write a proposal
3. Explore and validate the utility of the data
4. Write a professional report