

# Day One: What is Data Science?

Data Science Immersive

# Outline

- Introductions
- What is Data Science?
- What will we learn during the next 15 weeks?
- Explore the data science process.

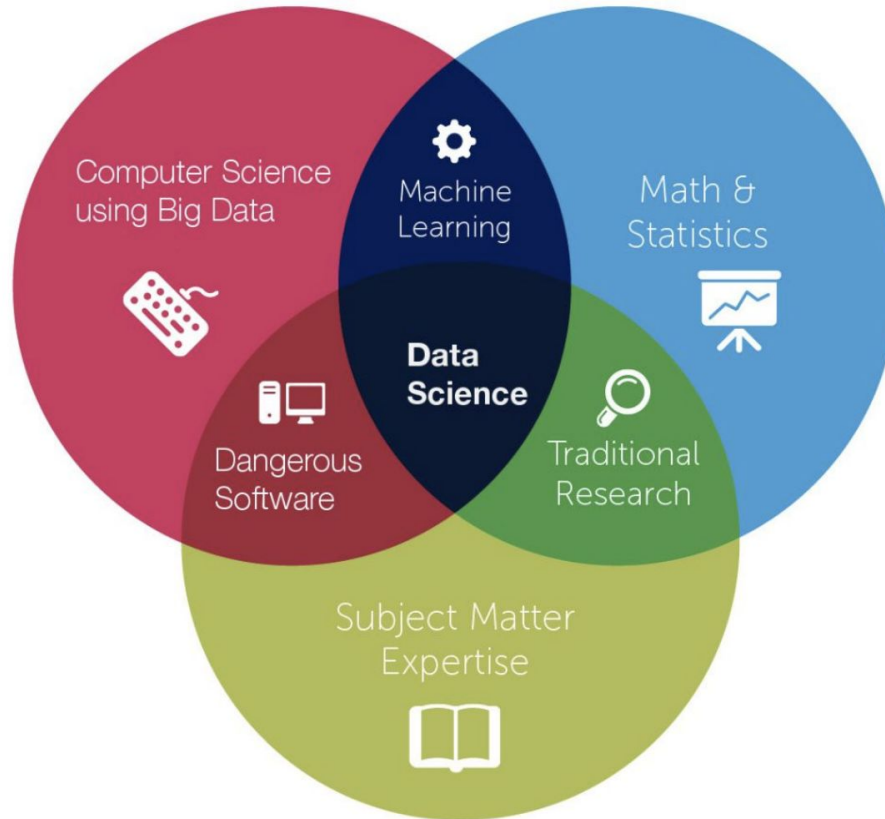
# Introductions

1. Name
2. What were you doing right before coming to Flatiron?
3. Fun fact!

# What is Data Science?

According to Wikipedia, the definition of data science is “an **interdisciplinary** field that uses scientific methods, processes, algorithms and systems to extract **knowledge and insights from various forms**, both structured and unstructured”

# The Classic Data Science Venn Diagram



# Defining Data Science

- “Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.”

# What kinds of questions can DS answer?

- How many minutes, on average, is an American person exposed to advertisement?
- Given the items you have purchased in the past, what items are you more likely to purchase in the future?
- Does the \$5 coupon or 25% off result in more returning customers for your favorite store?
- Is a given image a llama or duck?
- How should the government allocate educational resources?

# Data Science Process

- *Depends on what question you're trying to answer*





# Mod 1: Python and Data Engineering

- Introductory Python
- Object Oriented Programming
- Data Cleaning/Visualization
- Web scraping, APIs, and JSON
- SQL (and using SQL with Python)

# Mod 2: Statistics and Intro to Machine Learning

- Probability distributions
- Statistical Tests
- Introduction to machine learning with regression

# Mod 3: Diving Into Machine Learning

- Regression and classification models
  - Tree-based models
  - Tuning and evaluating models
- Calculus and linear algebra


# Mod 4: Advanced Topics

- NLP -Treating words as numbers
- Recommendation Engines
- Time Series
- Clustering
- Deep learning

# Weeks 13-15: Final Project

- One big long project - start to finish process
  - Data gathering
  - Research-motivated data processing and modeling
  - Presentable conclusions and/or model implementation
- Work independently
- A few optional advanced lectures

# Elements of the Course

- Learn.co as textbook
  - Readme's
  - Labs
- Lectures & presentations from Instructor and Coaches
- Pair Programming
- Blogging
- Assessments
- Projects & Science Fair
- Feelings Fridays
- Other community events 

# Tools We Use

- Python 3.7 and Anaconda
  - Make sure you can run Jupyter Notebook
  - Pip and Homebrew
- ~~Excel~~ - you'll never want to use it for data analysis again
- Slack - desktop app
  - Mobile app highly recommended (iOS & Android)
- Class calendar
- Github Account
- Learn Environment (learn-env)
- Zoom

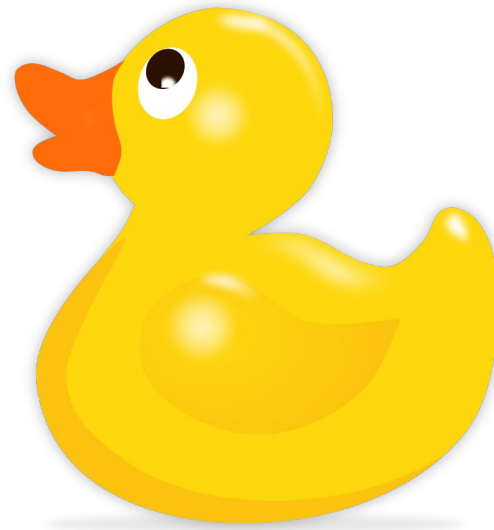
# Non-Technical Challenges

- Imposter Syndrome
- Too Much Material
- Jack of all trades, master of none
- Burnout
- Remote challenges



# Working Through Issues

- Google/Stackoverflow
- Phone a friend
- Rubber Ducking



# Remote Learning Additions

- **Mod 1 Schedule:**

<https://docs.google.com/spreadsheets/d/1Z2UHwkhPKQIP7Jk2t94RLj6-t7b-u6zGSvm5BS2Ollw/edit?usp=sharing>

- **1 on 1s:** Planned 15 minute meetings with a coach or instructor
- **Office Hours:** Zoom room where you can pop in and out with questions.

# Next Steps

- Conda install
- Be able to pull a Jupyter notebook from the command line
- Do not worry about “**Setting Up Virtual Environments**”
- Install an IDE. If you don't know what this is, install PyCharm
- Install Git on your computer.

# Next Steps

- Conda install
- Be able to pull a Jupyter notebook from the command line
- Do not worry about “**Setting Up Virtual Environments**”
- Install an IDE. If you don't know what this is, install PyCharm
- Install Git on your computer.

# Next Steps

- Labs and readings through Learn.co
- Github Lecture at 11AM

# Next Steps

*Interesting Data Science links:*

- [Battle of the Data Science Venn Diagrams](#)
- [10 Data Analytics Success Stories: A Closer Look](#)
- [Supervised vs. Unsupervised Learning](#)
- [Beyond Interactive: Notebook Innovation at Netflix](#)
- [The 25 Best Data Visualizations of 2018](#)

# The End



# Activity

- Talk to your neighbor and identify a topic that you would like to investigate for one of your projects during this bootcamp.
- Then brainstorm a question you would like to answer about that topic and lay out your plans or thought process to solve that problem.
- Use a whiteboard to draw a picture that represents that process. Think about the following in your graphic.
  - what goes in
  - what comes out
  - what happens in the middle



# Terminal 101

## Basic terminal commands:

- `pwd` → print working directory
- `mkdir` → make a directory
- `ls` → list of documents in a directory
- `cd` → change directory
- `cd ..` → goes back to previous directory
- `cd ~` → goes back to root directory
- `Mv` → move stuff around; change name
- `rm` → remove stuff
  - `rm -r directory`
- Opening jupyter notebook from terminal by typing command “Jupyter Notebook”