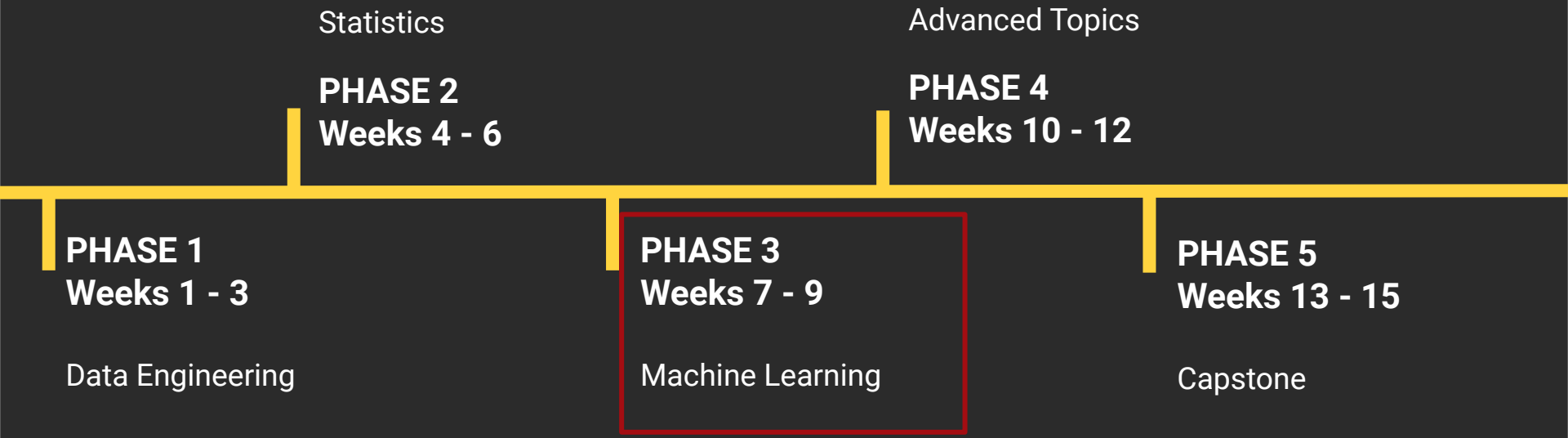

Phase 3 Expectations
Machine Learning

Agenda

1. Phase 3 Overview
2. Week 1: Math & Classification
3. Week 2: More Models!
4. Gating / Assessments
5. What Will Phase 3 Feel Like?

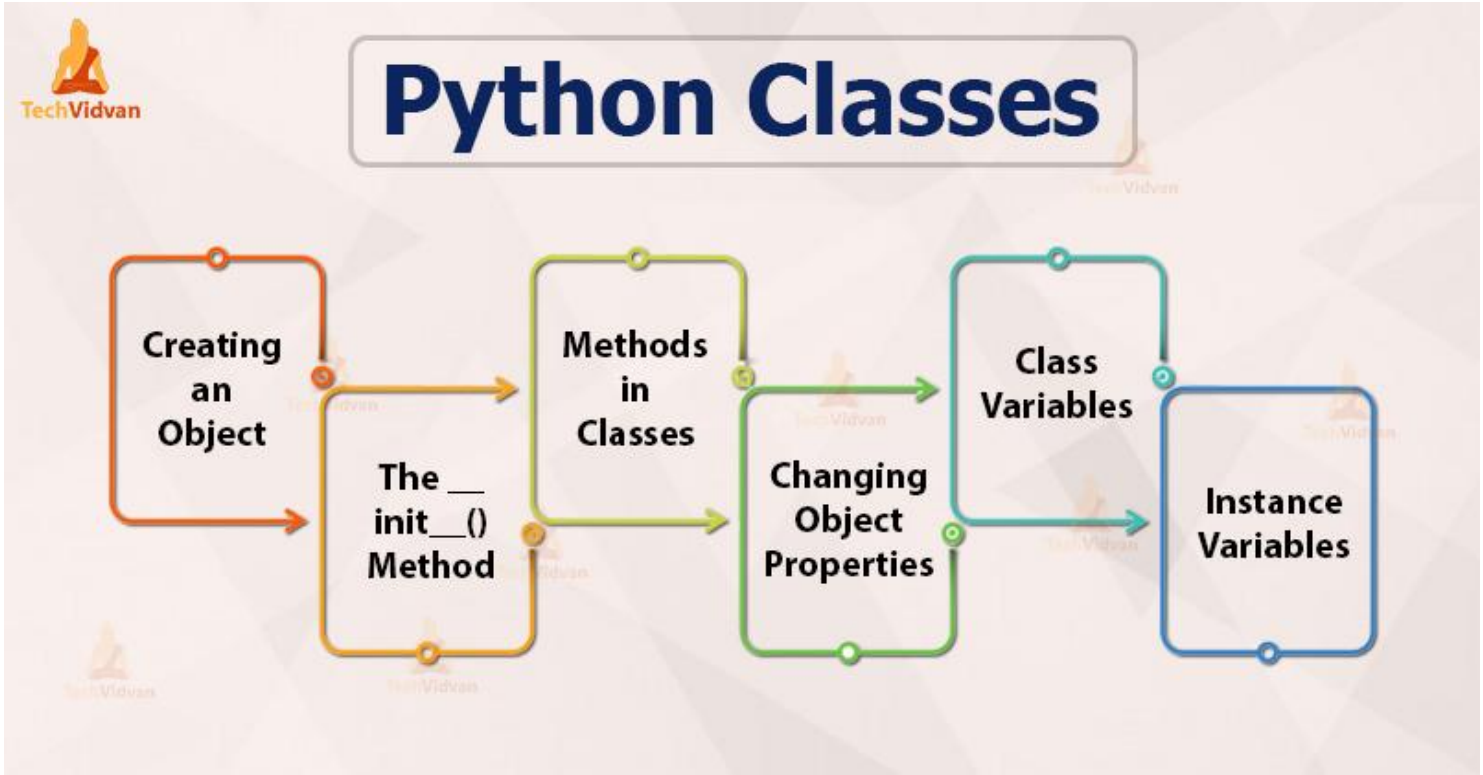
Overall Phase Timeline



Week 1

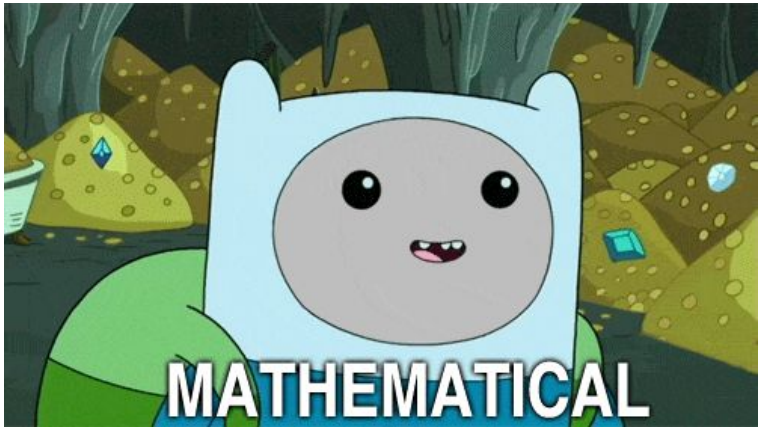
Programming, Math, Introduction to Classification

Object-Oriented Programming



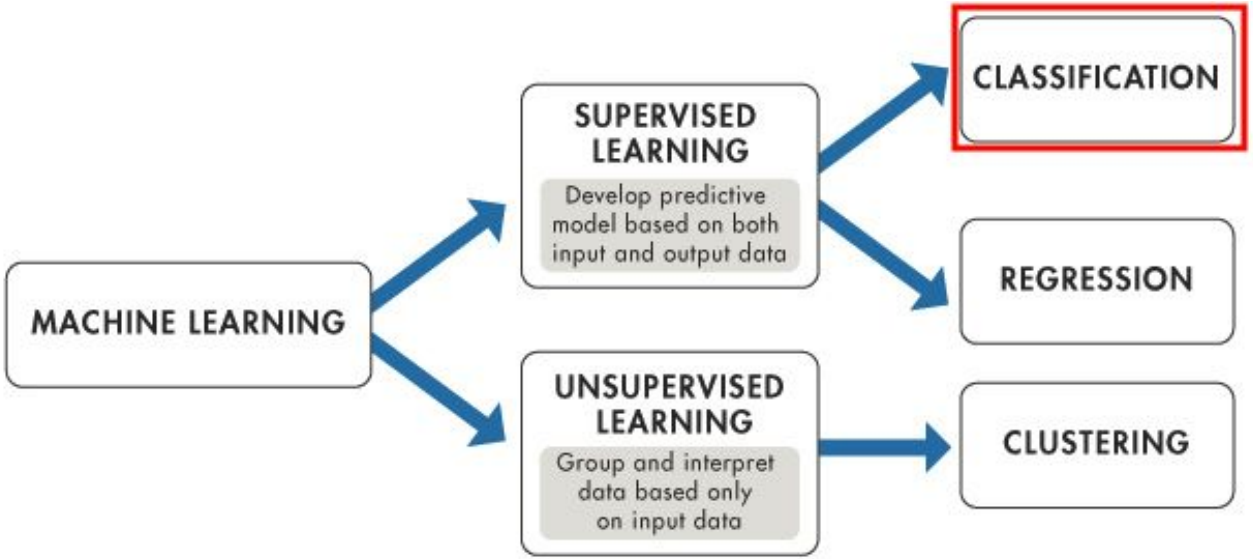
The Math Behind Data Science

- Linear Algebra
- Calculus, Cost Functions, Gradient Descent
- Focus on concepts and application!

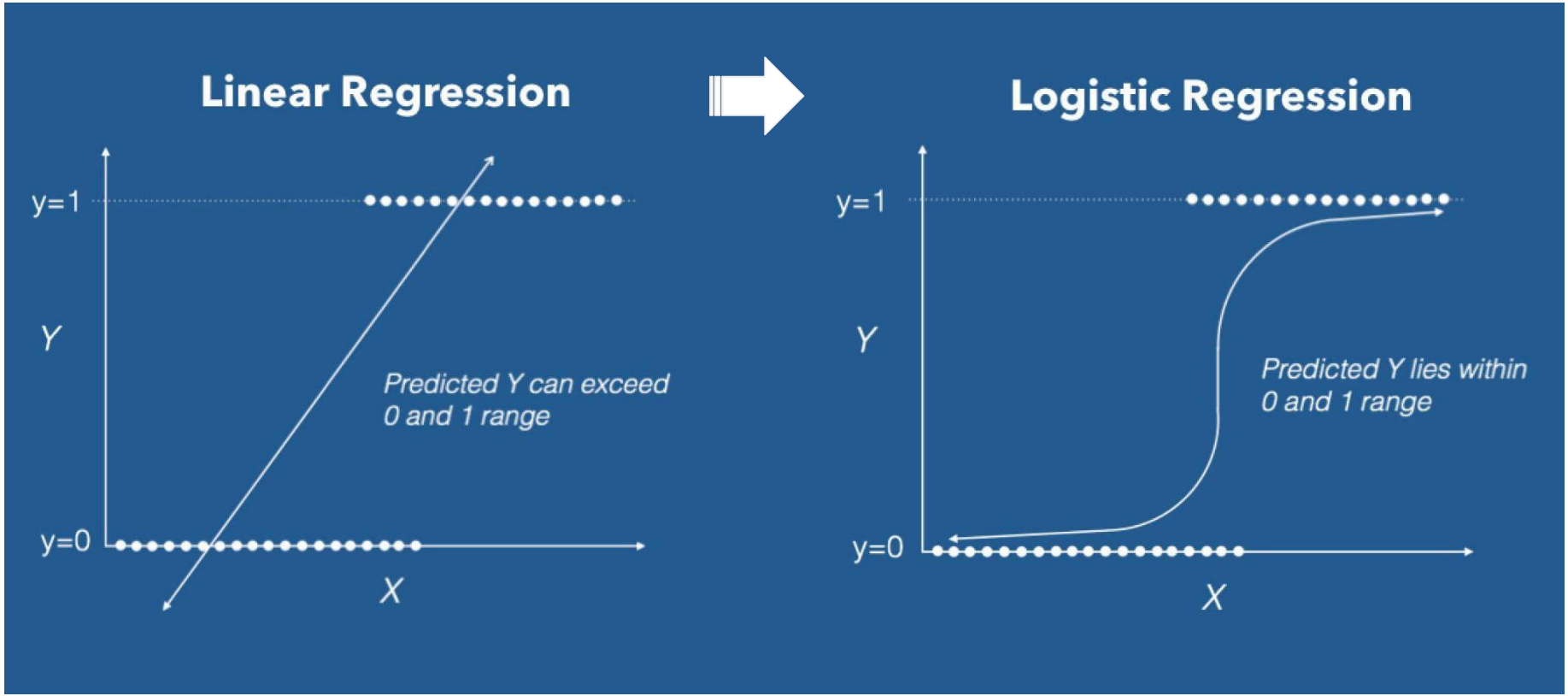


Classification

- Asking: Is it ____ or not?
- From continuous target to categorical target



Logistic Regression

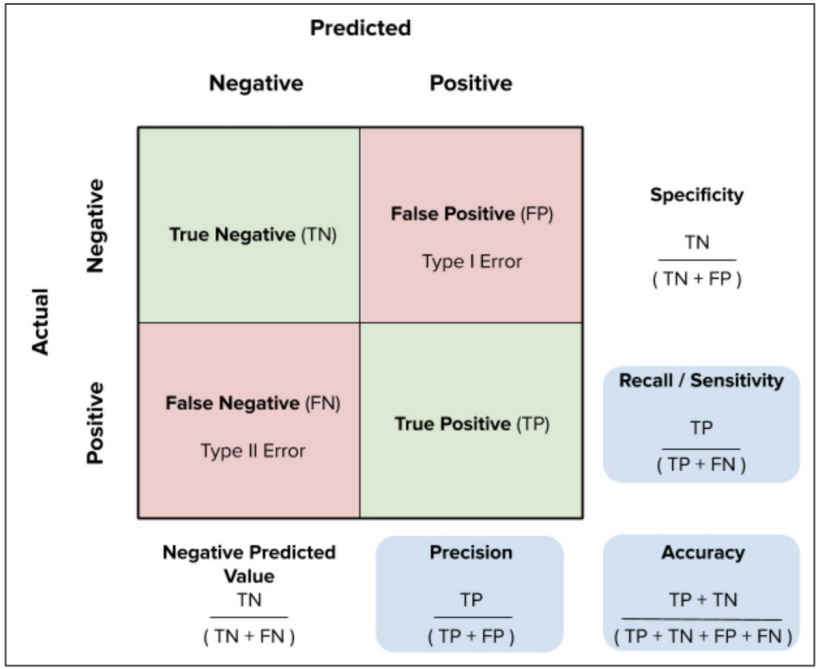


Week 2

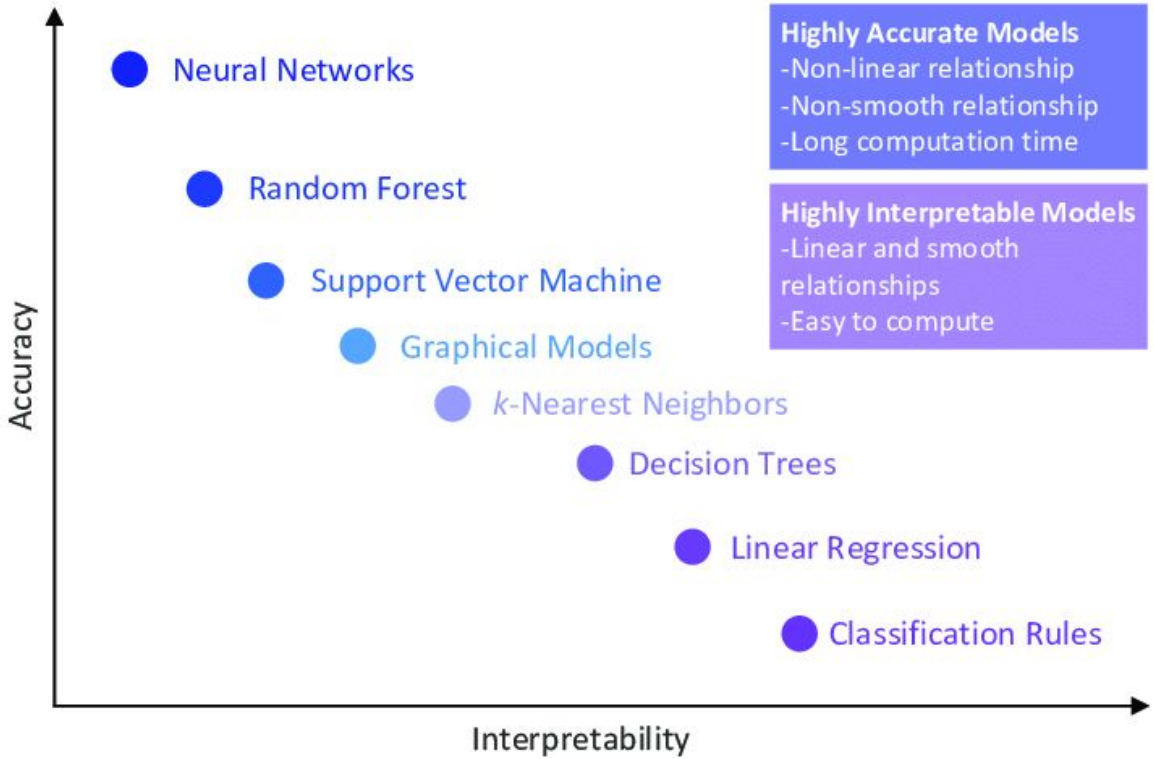
Classification Evaluation, More Models, Model Tuning

Classification Metrics

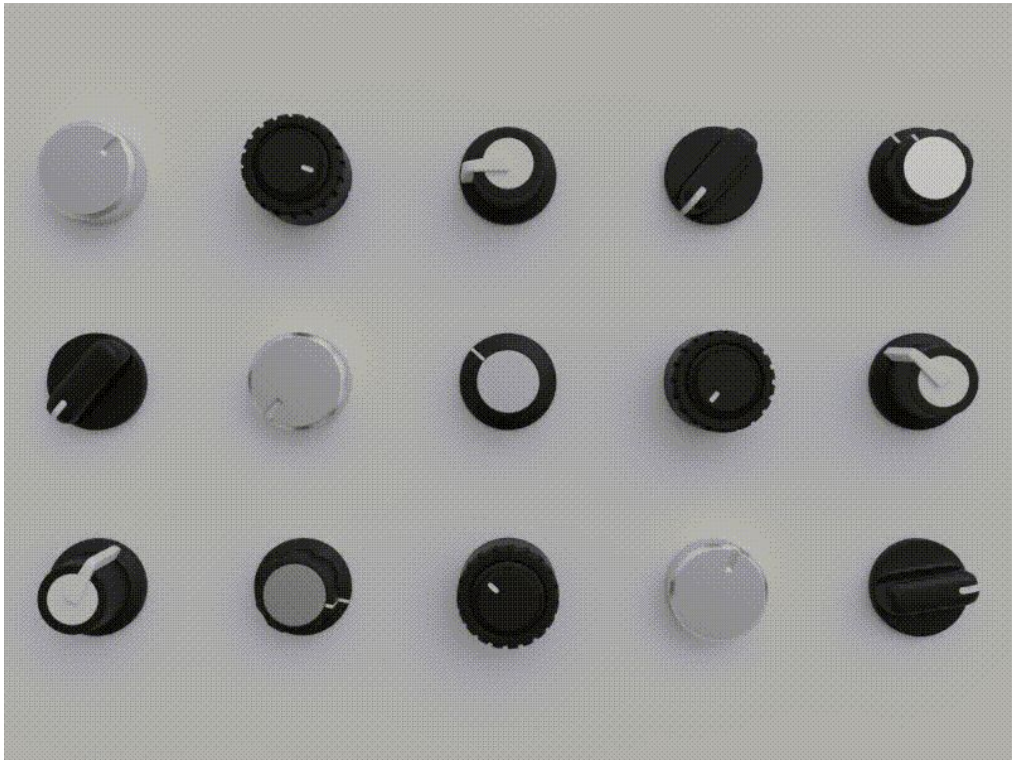
- New type of target - so, new metrics!



More Models!


Neural Networks
Random Forest
Support Vector Machine
Graphical Models
k-Nearest Neighbors
Decision Trees
Linear Regression
Classification Rules
Accuracy
Interpretability
Highly Accurate Models
-Non-linear relationship
-Non-smooth relationship
-Long computation time
Highly Interpretable Models
-Linear and smooth relationships
-Easy to compute

Model Tuning



Measuring Student Progress

1. **Blog Post**
    a. Due the **Tuesday** of the Second Week (Nov, 8th)
    b. Topic: Write a Tutorial (See next slide)
2. **Canvas & Code Challenge**
    a. Checkpoints
        i. Tuesday Nov, 1st: OOP checkpoint
        ii. Thursday Nov, 3rd: ML Fundamentals Checkpoint
        iii. Monday Nov, 7th: Logistic Regression Checkpoint
        iv. Wednesday Nov, 9th: Decision Trees Checkpoint
    b. Code Challenge
        i. **Thursday Nov, 10th**

Measuring Student Progress: Blog #3

Write a tutorial (with a data set and code sample) on something that you think might be interesting to other people taking the course. It can be a topic we didn't cover at all, or can just go deeper into a topic that we did cover. Your tutorial can use the same tool/library from the previous blog post, or you can choose something new. Potential elements to include:

I. An introduction explaining why a data scientist would want to do what your tutorial does
II. A section explaining the data set you're working with (what are the features? if there is a target, what is it?)
III. A section explaining the libraries you're working with (ideally with links to documentation and/or tutorials for someone just getting started)
IV. Well-commented code
V. A conclusion section that recaps what you did and why

What can you expect to ***feel*** when going through Phase 3?

Feelings at the Start of Phase 3

Excited:
Let’s tackle Machine Learning!

Overwhelmed:
Many new DS tools to learn

Feelings at the End of Phase 3

Stressed:
Starting to think about capstone and end of course

Powerful:
Armed with modeling tools to attack many different problems

Capstone Project

- Sneaking up!
- What can you do now?
  - Explore data!
    - UCI Machine Learning Datasets Repository
    - Kaggle Datasets
    - Awesome Datasets Repo on Github
    - New York City Open Data Portal
    - Inside AirBNB
    - Data is Plural


LET'S
DO THIS!
KUNG FU PANDA © 2019 DREAMWORKS ANIMATION LLC. ALL RIGHTS RESERVED.