GENERAL ASSEMBLY

INTRODUCTION TO MODELING

Matt Brems

General Assembly, DSI+

DATA SCIENCE PROCESS

1. Define problem.
2. Gather data.
3. Explore data.
4. Model with data.
5. Evaluate model.
6. Answer problem.

MODELING

- Modeling is something that we naturally do.
  - How long does it take you to get to work?

MODELING

- Modeling is something that we naturally do.
  - How long does it take you to get to work?

- A **model** is a simplification of reality.

- How do we simplify?
  - Taking into account only really important factors.

  - Making assumptions about how things behave.

MODELING

"Essentially, all models are wrong, but some are useful."

– George Box, 1987

MODELING

MACHINE LEARNING ALGORITHMS

- **Machine learning** is a term we use to describe getting computers (machines) to learn without needing to be explicitly programmed.

- There are many different machine learning algorithms we'll cover in the class - from linear regression to neural networks!

MACHINE LEARNING ALGORITHMS

TERMINOLOGY

- $\boldsymbol{X}$: our data, the independent/explanatory variables we use to predict $\boldsymbol{Y}$.

- $\boldsymbol{Y}$: our data, the dependent variable we want to predict.

- $\boldsymbol{\widehat{Y}}$: our predicted values of $\boldsymbol{Y}$.