

Predictive Analytics Foundations

Lecture 1

Lecture 1



- **Intro**
- What is predictive analytics?
- Data Science Lifecycle
- What will you learn in this class?
- Demo



Introduction

Meet with your instructor

Lecture 1



- Intro
- **What is predictive analytics?**
- Data Science Lifecycle
- What will you learn in this class?
- Demo

Technology Trends

2020s ● ?

2010s ● Data Industry

- Collect and sell information

2000s ● Internet Industry

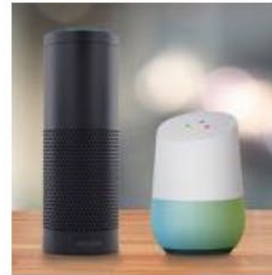
- Online retailers and services

1990s ● Software Industry

- Sold computer software

1980s ● Hardware Industry

- Sold computers



Predictive Analytics

- **Exploration**
 - Identifying patterns in data
 - Uses visualizations
- **Inference**
 - Using data to draw reliable conclusions about the world
 - Uses statistics
- **Prediction**
 - Making informed guesses about unobserved data
 - Uses machine learning

Some (broad) questions we might try to answer with predictive analytics

- In which markets should we focus our advertising campaign?
- Should I send my kids to daycare?
- Is the world getting better or worse?
- What areas of the world are at higher risks for climate change impact in 10 years? 20?
- What should we eat to avoid dying early of heart disease?
- Do immigrants from poor countries have a positive or negative impact on the economy?

Why Predictive Analytics Matters

Regularly Eating Chocolate Is Linked to 8 Percent Lower Heart Attack Risk



By [Lisa Rapaport](#)

Reviewed: July 23, 2020

Report on a July 2020 article in
the European Journal of
Preventive Cardiology



<https://journals.sagepub.com/doi/full/10.1177/2047487320936787>

<https://www.everydayhealth.com/diet-nutrition/eating-chocolate-regularly-linked-to-lower-heart-attack-risk/>

Observation

- **individuals**, study subjects, participants, units
 - *336,289 US, Swedish, and Australian adults in several studies*
- **treatment**
 - *chocolate consumption*
- **outcome**
 - *heart disease*

The First Question

Is there **any relation** between chocolate consumption and heart disease?

- **association**
 - any relation
 - link

Answer: Yes, because those who ate chocolate had less heart disease.

Other headlines about the same article:

A Stronger Link?

Is eating chocolate heart-healthy? Study says 'yes'

August 26, 2020 No Comments    

Family Safety and Health, National Safety Council

Chocolate is good for the heart

22 Jul 2020

European Society of Cardiology Press Release

<https://www.escardio.org/The-ESC/Press-Office/Press-releases/Chocolate-is-good-for-the-heart>

<https://www.safetyandhealthmagazine.com/articles/20257-is-eating-chocolate-heart-healthy-study-says-yes>

The Next Question

Does chocolate consumption **lead to** a reduction in heart disease?

- **causality**

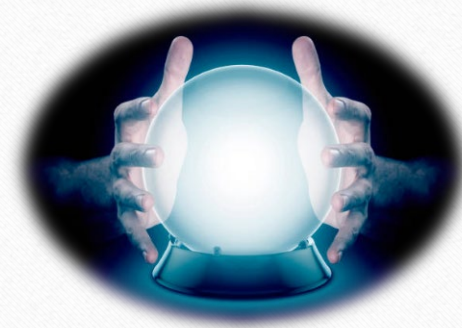
This question is often harder to answer.

“Dr. Alice Lichtenstein, an American Heart Association volunteer and professor of nutrition science and policy at Tufts University, was more skeptical of the findings.”

Market Watch

Association ? Causation?

Prediction?

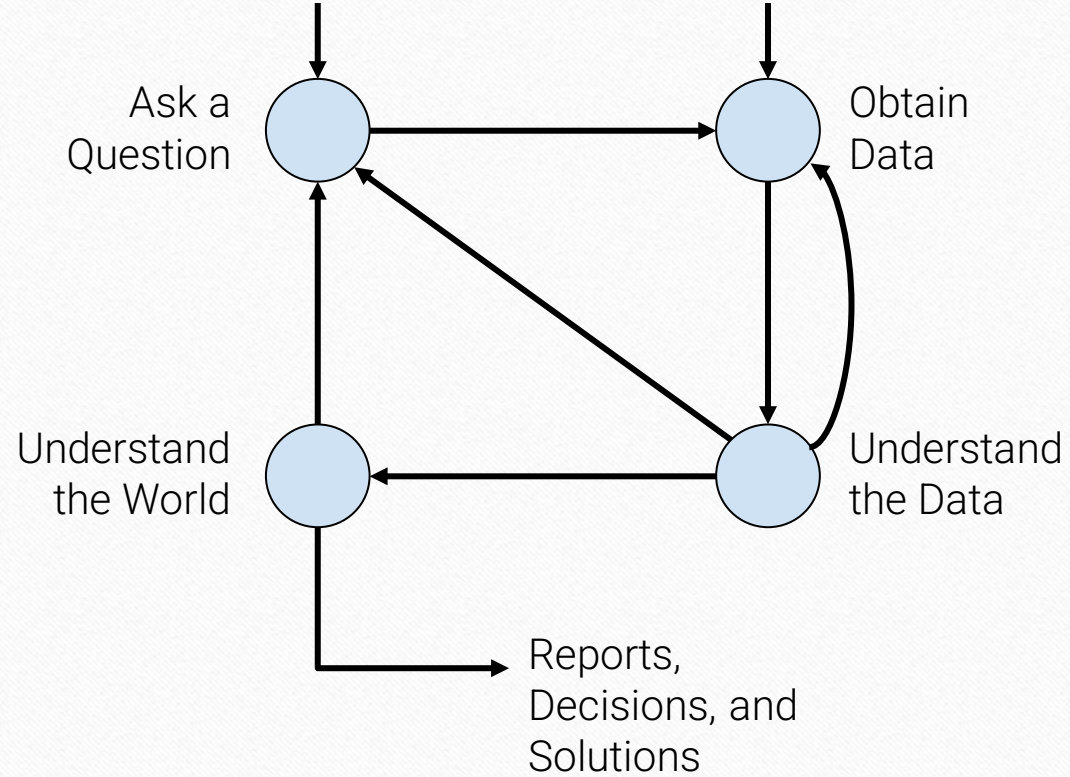


Lecture 1



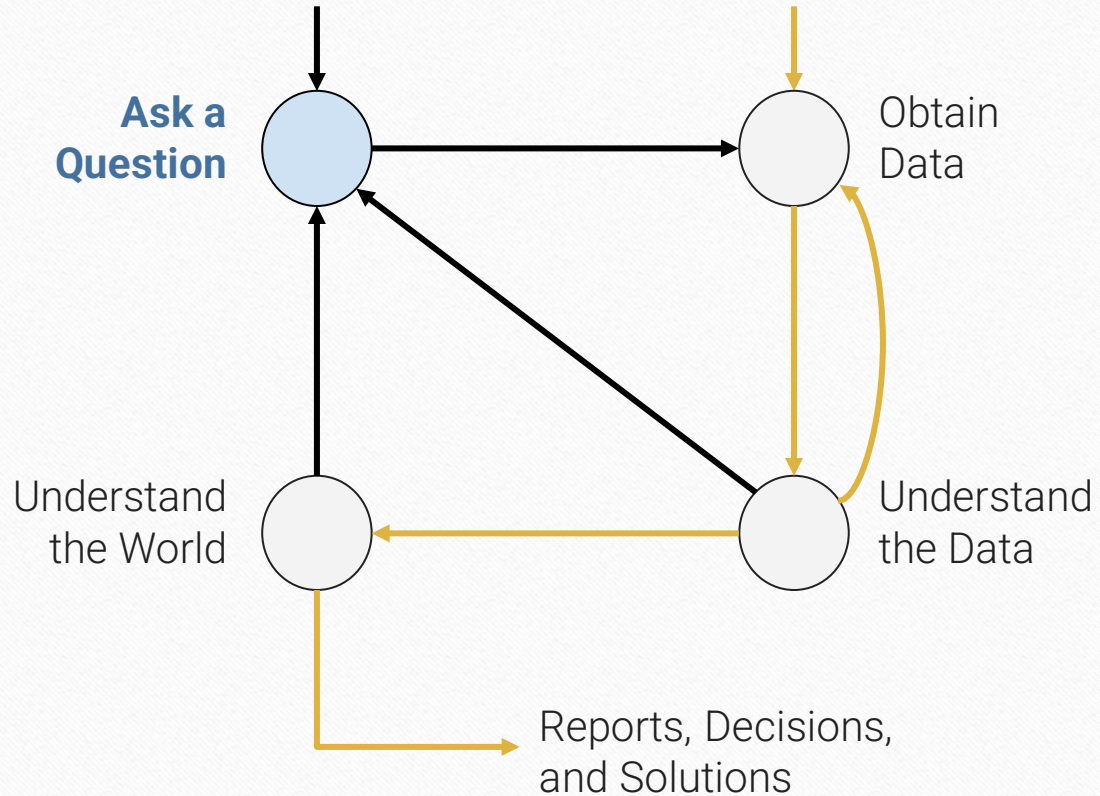
- Intro
- What is predictive analytics?
- **Data Science Lifecycle**
- What will you learn in this class?
- Demo

Data Science Lifecycle



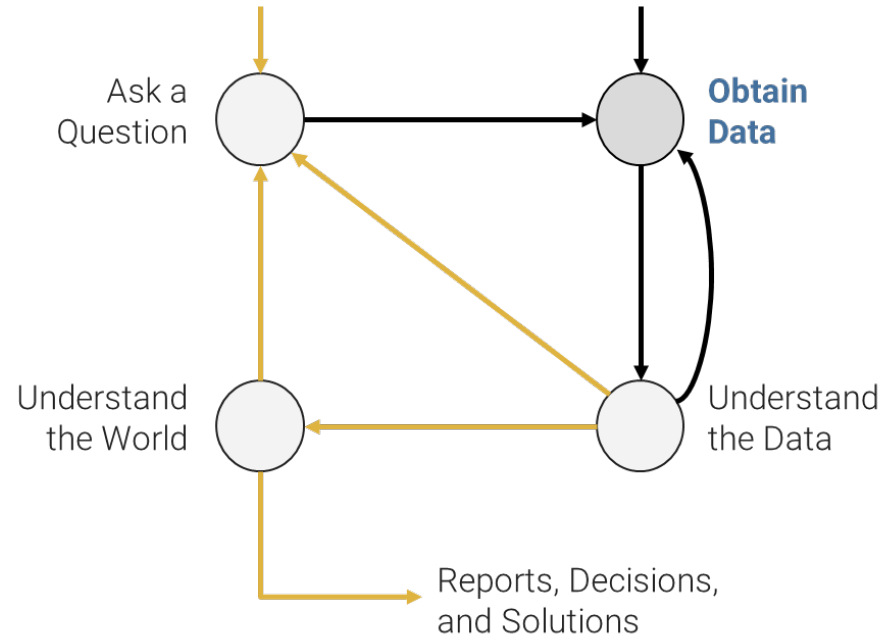
1. Question/Problem Formulation

- What do we want to know?
- What problems are we trying to solve?
- What are the hypotheses we want to test?
- What are our metrics for success?



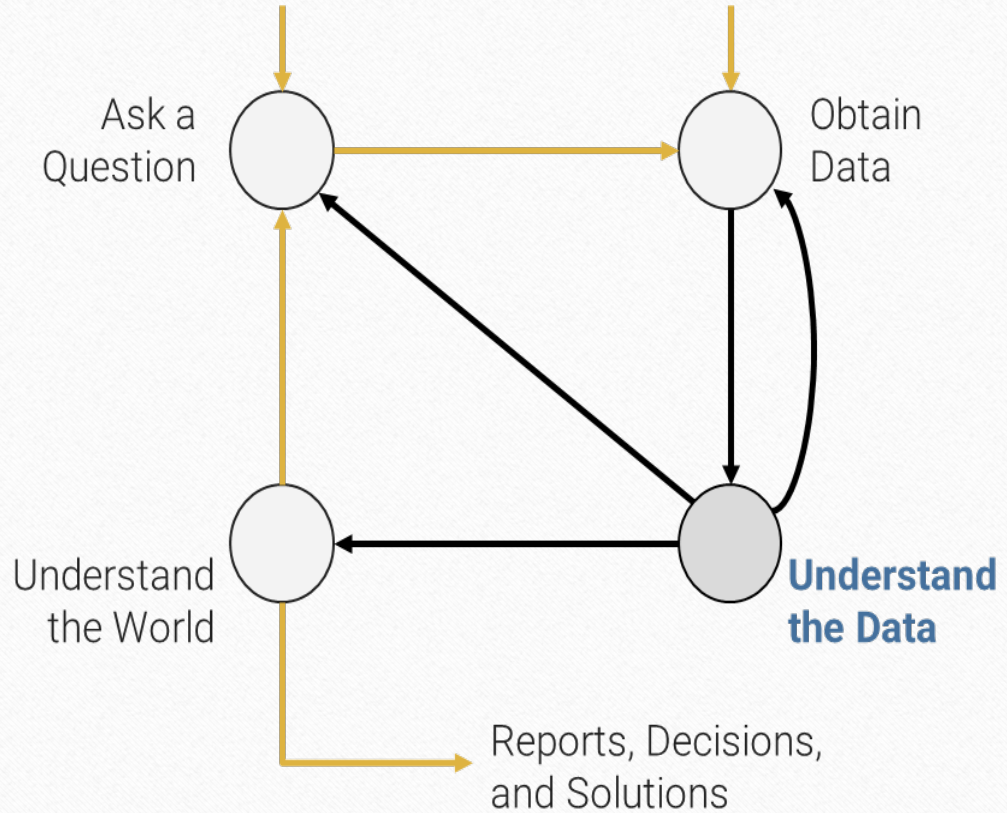
2. Data Acquisition and Cleaning

- What data do we have and what data do we need?
- How will we sample more data?
- Is our data representative of the population we want to study?



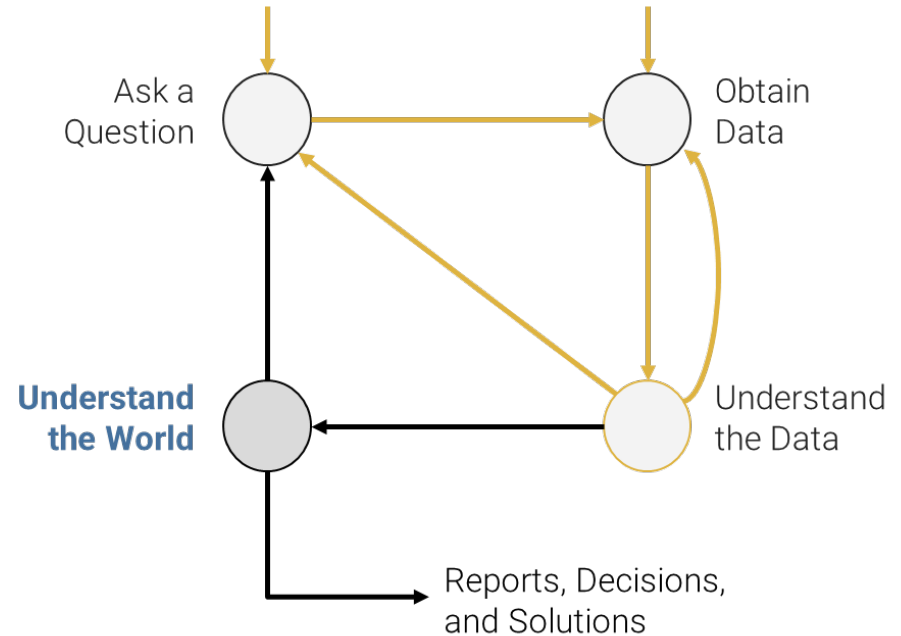
3. Exploratory Data Analysis & Visualization

- How is our data organized and what does it contain?
- Do we already have relevant data?
- What are the biases, anomalies, or other issues with the data?
- How do we transform the data to enable effective analysis?



4. Prediction and Inference

- What does the data say about the world?
- Does it answer our questions or accurately solve the problem?
- How robust are our conclusions and can we trust the predictions?



Lecture 1



- Intro
- What is predictive analytics?
- Data Science Lifecycle
- **What will you learn in this class?**
- Demo

Course Content



Textbook

Assessments

- Lab exercises
- Quizzes
- Midterm
- Final exam

Collaboration

Asking questions is greatly encouraged

- Discuss questions with each other (except exams)
- Submit lab assignments individually

The Limits of collaboration

- Don't share solutions with each other
- Copying or other dishonesty will result in severe penalties

Lecture 1



- Intro
- What is predictive analytics?
- Data Science Lifecycle
- What will you learn in this class?
- **Demo**