

Introduction to Data Science

What is Data Science?

Joanna Bieri DATA101

Info

Important Information

- Email: joanna_bieri@redlands.edu
- Office Hours take place in Duke 209 – Office Hours Schedule
- Class Website
- Syllabus

What is data science?

What is Data?

- In our everyday life, we are constantly surrounded by data.
 - Personal Information
 - Words and Text
 - Scientific Studies
 - Surveys
- What are some examples of data?

What is Data Science?

On Wikipedia, Data Science is defined as a scientific field that uses scientific methods to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains.

What is Data Science?

- The main goal of data science is to **extract knowledge** from data.
- To **understand** data
- Find hidden relationships and build a **model**

What is Data Science?

- Data science uses **scientific methods**, such as probability and statistics.
- More than just a fancy name for statistics.
- Uses math, computer science, writing, art, communication.

What is Data Science?

- Obtained knowledge should be applied to produce some **actionable insights**
 - Communicate your knowledge.
 - Enact Change

What is Data Science?

- **Application domain** is an important concept!
- Data scientists often need at least some degree of expertise in the problem domain.
 - finance
 - medicine
 - marketing
- What are some of your interests?
- Where might you apply data science?

What are some Related Fields?

Since data is pervasive, data science itself is also a broad field, touching many other disciplines.

- **Databases**
- **Big Data**
- **Machine Learning**
- **Artificial Intelligence**
- **Visualization**

Types of Data

What are the types of data?

- Structured - numbers in a table
- Unstructured - a collection of files to be processed
- Semi-structured - somewhere in between

Types of Data - Examples

Structured	Semi-structured	Unstructured
List of people with their phone numbers	Wikipedia pages with links	Text of Encyclopedia Britannica

Types of Data - Examples

Structured	Semi-structured	Unstructured
Temperature in all rooms of a building at every minute for the last 20 years	Collection of scientific papers in JSON format with authors, date of publication, and abstract	File share with corporate documents

Types of Data - Examples

Structured	Semi-structured	Unstructured
Data for age and gender of all people entering the building	Internet pages	Raw video feed from surveillance camera

Types of Data - Examples

Structured	Semi-structured	Unstructured
??	??	??

Where can you get data?

- **Structured**

- **Internet of Things (IoT)**, including data from different sensors, such as temperature or pressure sensors, provides a lot of useful data. For example, if an office building is equipped with IoT sensors, we can automatically control heating and lighting in order to minimize costs.
- **Surveys** that we ask users to complete after a purchase, or after visiting a web site.
- **Analysis** of behavior can, for example, help us understand how deeply a user goes into a site, and what is the typical reason for leaving the site.

Where can you get data?

- **Unstructured**

- **Texts** can be a rich source of insights, such as an overall sentiment score, or extracting keywords and semantic meaning.
- **Images or Video.** A video from a surveillance camera can be used to estimate traffic on the road, and inform people about potential traffic jams.
- **Web server Logs** can be used to understand which pages of our site are most often visited, and for how long.

Where can you get data?

- **Semi-structured**
 - **Social Network graphs** can be great sources of data about user personalities and potential effectiveness in spreading information around.
 - When we have a bunch of photographs from a party, we can try to extract **Group Dynamics** data by building a graph of people taking pictures with each other.

Why is Data Science so AWESOME?

There are lots of reasons:

- You can tell an effective story and make a point about something you care about.
- You can make really beautiful visualizations.
- You can understand the world and answer questions about the world.
- You can get a job!

Great examples of impressive Data Science Visualizations and Projects

The World's Population at 8 Billion

Great examples of impressive Data Science Visualizations and Projects

The History of Pandemics

Great examples of impressive Data Science Visualizations and Projects

200 Years of Immigration to the US

Great examples of impressive Data Science Visualizations and Projects

World's Most Visited Websites