

Unit 1: What is data science?

What is data?

Lesson 4

Derek Ruths

Overview of unit

- What is data science?
- **What is data?**
- Data collection and curation
- What is data analytics?
 - How does data analytics fit into data science?
- What is machine learning?
 - How does machine learning fit into data science?

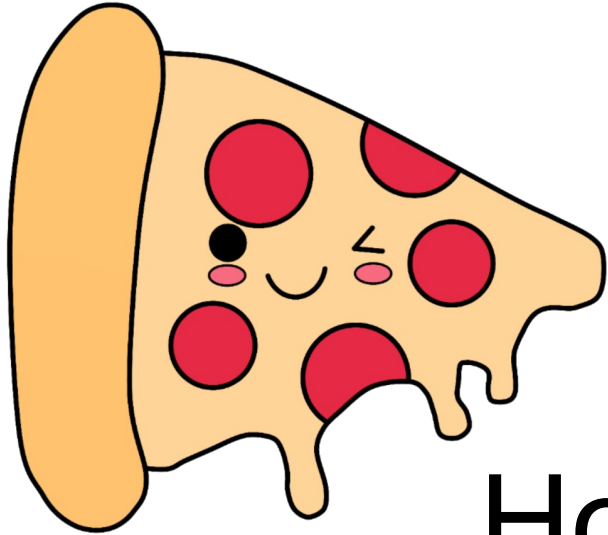
Lesson overview

Objectives

- Understand how quantification relates to data
- Recognize and propose proxies

Outline

- The essence of data
- Features
- Proxies



How much do people like pizza?

The essence of data

Data is a collection of recorded observations.

The essence of data

Quantitative data is data that can be represented as numbers.

Features

columns
of dataframe

A feature is a quantitative (possibly inferred)
attribute of a data.

eg Start with text
end with #

Proxies

Not all inferred features
are proxies
but some are

A proxy is a feature that approximates a particular thing we want to quantify.

eg happiness

Lesson wrap-up

Takeaways

- Data, quantitative data, features, proxies
- Probably most people like pizza

Up next

- Data science project phases in detail