

Introduction to Data Science

Lesson 1: Introduction to Data Science

Activity 2: Exploring a large data set

Open the spreadsheet for the data set. There is also a video provided in the Large Data Set section of the course that explains how the spreadsheet is structured.

Answer the following questions:

1. The sheet or sheets in the data set are **rectangular**: they contain columns of data for the **fields** (properties or attributes) and rows for the **records** (each row represents a single instance that the data is being recorded for).
 - a. How many fields are there in each sheet?
 - b. How many records are there in each sheet?

2. Which of the fields are numerical (showing measures or quantities) and which are categorical (text or options from a list of items)?

3. Why has this data been stored?

Consider who might want this data and what could they do with it.

4. The video contains some questions that you could answer using this data. What other questions could you answer with this data?

Typical questions are: "Is there a difference between ... for different types of ...?" and "Is there a link between ... and ...?"

5. This data is a subset of a much bigger data set. How big is the data set it is taken from? How many fields and records could be involved?