

Module overview: Preparing data

Getting set up for Preparing data

Overview

The tool we will be using in **Preparing data**:



In this module, we focus on **developing essential data** and **spreadsheet skills** in **Google Sheets**.

Before we can start our exploration of data, we will need to do the following:

Log into our existing **Google account**.

OR

Create a new **Google account**.

Access the data for the **specific** walk-throughs and learning activities.

OR

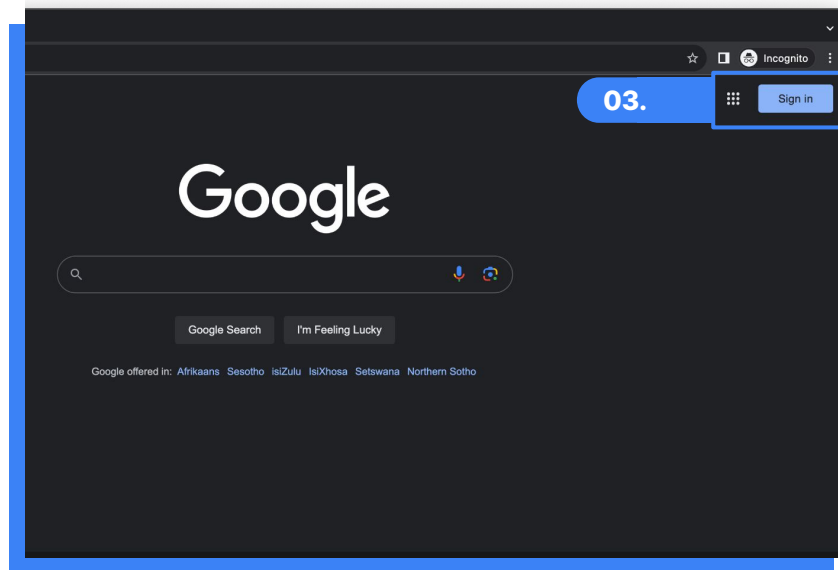
Download and **access all data** for Preparing data.

Google account

In order to use Google Sheets, we will need to have a Google account since our spreadsheets will be created and edited in the Google Drive linked to this account.

Logging into an **existing Google account**:

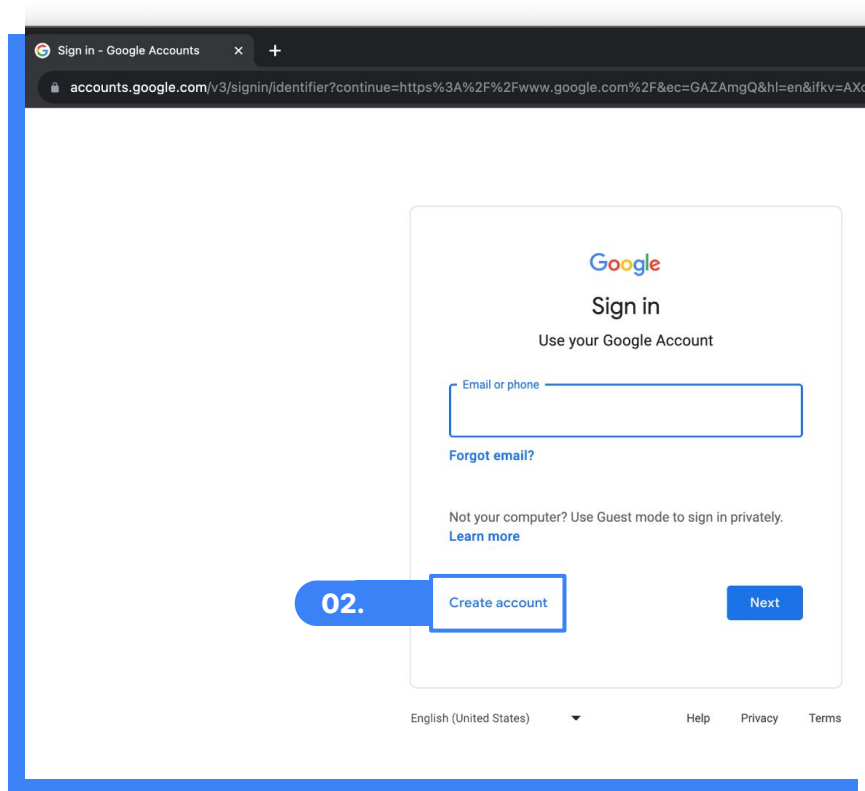
- 01.** Open a web browser (we recommend using Chrome).
- 02.** Navigate to Google <https://www.google.com/>
- 03.** Click on Sign in.
- 04.** Add your email address or phone number.
- 05.** Enter your password.
- 06.** If you have 2-factor authentication set up, follow the instructions to log in.
- 07.** You'll be redirected to the main Google search interface upon successful login.



Google account

Creating a new free **Google account**:

01. Follow steps 01. to 03. on *logging into an existing Google account*.
02. Click on Create account.
03. Follow the prompts to create a new account.

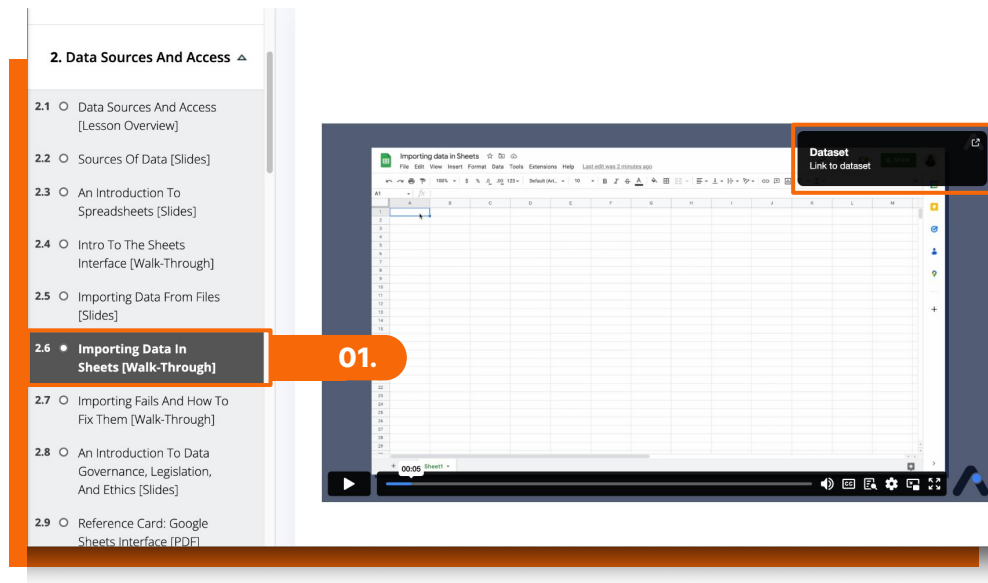


Accessing the data

There are multiple ways to access the data we need for this module, but all of the methods will require you to add the CSV and XLSX files to our Google Drive.

Accessing a **ZIP**, **CSV**, or **XLSX** (spreadsheet) file for a **specific walk-through**:

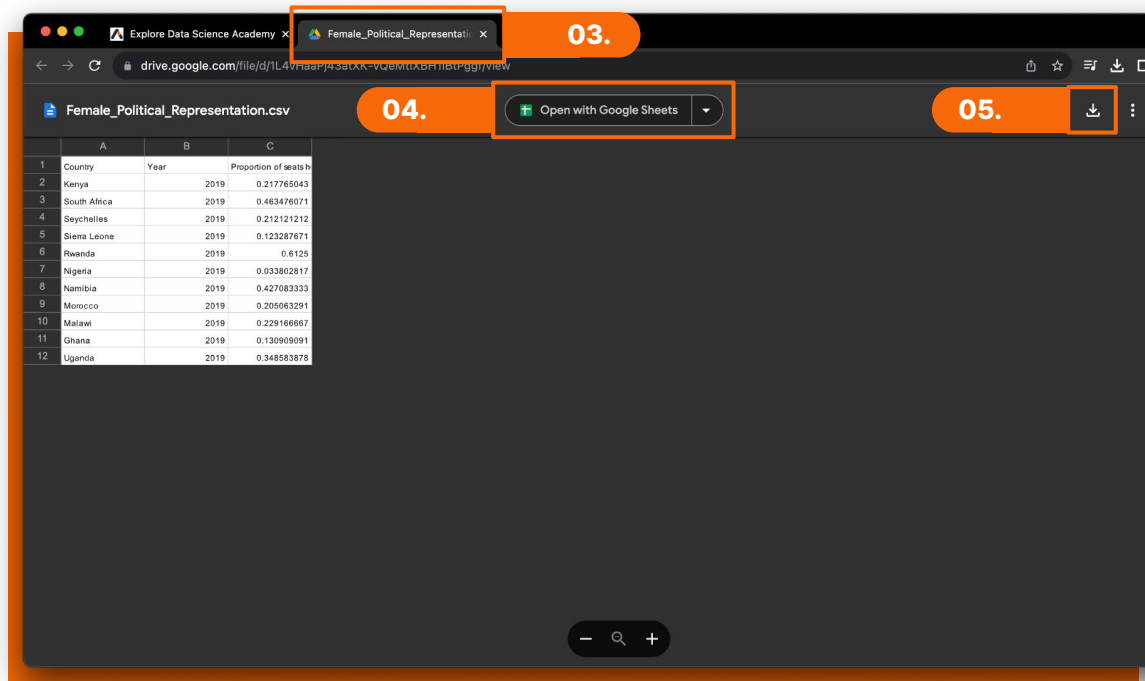
- 01.** On Athena, navigate to our piece of learning material that includes [Walk-through] in the title.
- 02.** Within the first few seconds of playing the video, a pop-up will appear in the top-right corner of the video frame.



Accessing the data

Downloading or opening **CSV** files:

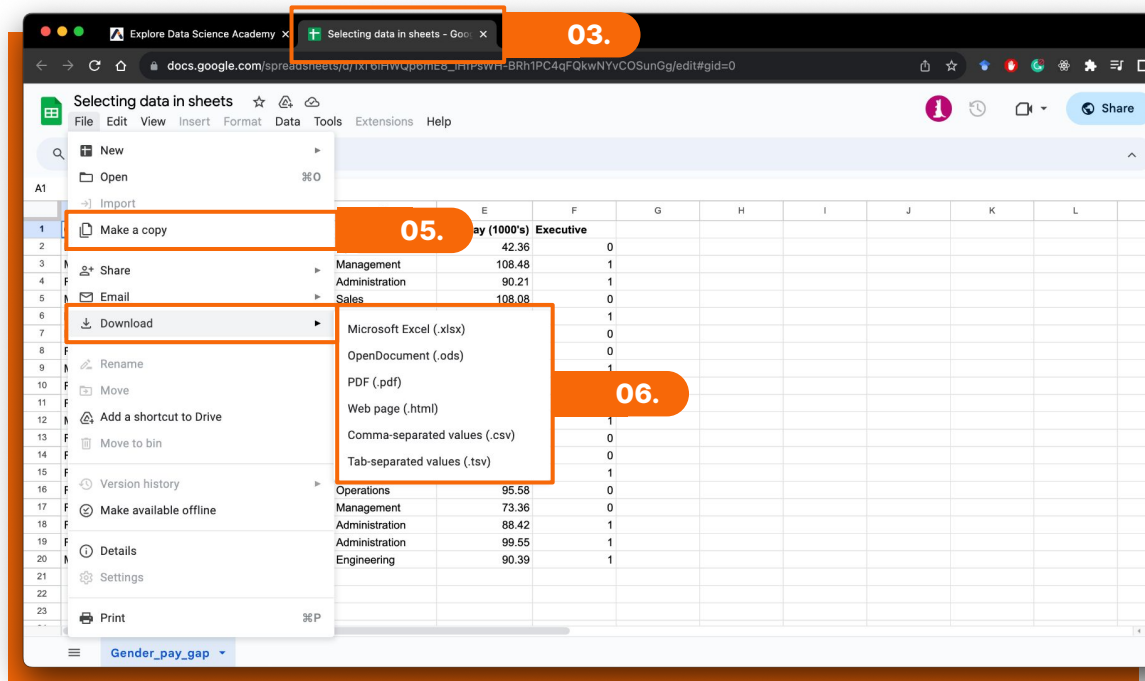
03. Clicking on the pop-up will open the file in a new tab.
04. We can open the file in Google Sheets by clicking on the recommended action button Open with Google Sheets.
Note: This action will automatically create a new spreadsheet file in our Google Drive.
05. We can also download the file to our local machine by clicking on the download icon. The file will be stored in our Downloads folder on our local machine.



Accessing the data

Downloading or opening **XLSX** (spreadsheet) files:

- 03. Clicking on the pop-up will open the spreadsheet file in Google Sheets in a new tab.
- 04. We only have View-only access to this file.
- 05. We can make a copy of the file and save it on our Google Drive.
- 06. We can download the spreadsheet file to our local machine as a XLSX, CSV, etc.



Accessing the data

Downloading the **entire module's datasets** and **spreadsheets** as a ZIP archive:

01. We navigate to this train, Getting set up for Preparing data [Slides].
02. Next to the description of this train on Athena, we can click on the Download Additional Files button to download the module's required datasets and spreadsheets.
03. We find the ZIP archive in the Downloads folder of our local machine. Double-click or right-click to unzip the archive.

The screenshot displays the Athena interface for the 'Preparing data' sprint. The sidebar on the left lists the following items:

- Sprint: Preparing data
- 1. Module Overview: Preparing Data
- 1.1 • Getting Set Up For Preparing Data [Slides] (highlighted with an orange box and labeled 01.)
- 1.2 ○ Integrated Project: Access To Drinking Water (Overview) [Slides]
- 2. Data Sources And Access
- 3. An Introduction To Using Data
- 4. Data Aggregations And Descriptive Statistics

The main content area shows the details for the 'Getting set up for Preparing data [Slides]' train. It includes a description: 'In this train, we look at what we need to set up to get started, including logging into a Google account or creating one, and how to start developing our data and spreadsheet skills. Download datasets we'll be using throughout this module from the Download Additional Files button.' Below the description are two buttons: 'Download Train' and 'Download Additional Files' (highlighted with an orange box and labeled 02.).

At the bottom of the interface is a large image of the Explore AI Academy logo.

Navigating Google Drive

Google Drive is a cloud-based file storage service that allows us to store, access, and manage files in real-time. Our Google Sheet files are saved on Drive, which means any updates we make automatically get synced and saved in the cloud.

01. Open a web browser (we recommend using Chrome).
02. Navigate to Google Drive <https://drive.google.com/>
03. If we copied a spreadsheet file in order to access a specific walk-through dataset, we should see it somewhere in our Drive, depending on where we copied the file to.
04. We can also use the search bar to find a specific file.

The screenshot shows the Google Drive web interface. At the top, there is a search bar labeled 'Search in Drive' which is highlighted with an orange box and labeled '04.'. Below the search bar, the 'My Drive' section is visible. On the left sidebar, the 'My Drive' option is selected. The main area displays a list of files. The file 'Selecting data in sheets' is highlighted with a blue box and labeled '03.'. The file list includes columns for Name, Owner, Last modified, and File size.

Name	Owner	Last modified	File size
Folder	me	16 Jan 2020	—
File	me	22 Aug 2022	1 KB
File	me	31 Oct 2022	7.6 MB
Selecting data in sheets	me	16:12 me	1 KB
File	me	22 Sept 2021	8.9 MB
File	me	23 Mar 2023	20.4 MB
File	me	22 Sept 2021	10 KB
File	me	28 Apr 2022	75 KB

Navigating Google Drive

Uploading and accessing the entire module's datasets and spreadsheet files:

01. Clicking on + New in the left panel of Google Drive, we select Folder upload.
02. Our local computer file system will open in a window. We can navigate to the unzipped folder we downloaded earlier and select upload.
03. We will see a pop-up in the bottom-right corner that provides the status of the upload.

