

✔ **Congratulations! You passed!**

Grade received **100%** To pass 100% or higher

[Go to next item](#)

## Hands-On Activity: Kaggle datasets

Total points 2

1.



1 / 1 point

### Activity overview

In the last activity, you got set up on Kaggle and explored the Notebooks feature. In this activity, we will work with a different feature of the Kaggle platform: datasets.

Kaggle has tens of thousands of datasets that are available for public use. Anyone can upload a dataset to Kaggle. If they choose to make it public, other Kagglers can use that dataset to create their own projects.

First, you'll take a tour of a specific dataset. Then, you'll have a chance to choose your own datasets to work with. Finally, you'll use what you've learned in this module to determine the kind of data in your datasets, and whether the data is biased or unbiased.

By the time you complete this activity, you will be able to use many of the helpful features Kaggle has to offer. This will enable you to find data for projects and engage with the data community, which is important for developing skills and networking in your career as a data analyst.

### Explore Kaggle datasets

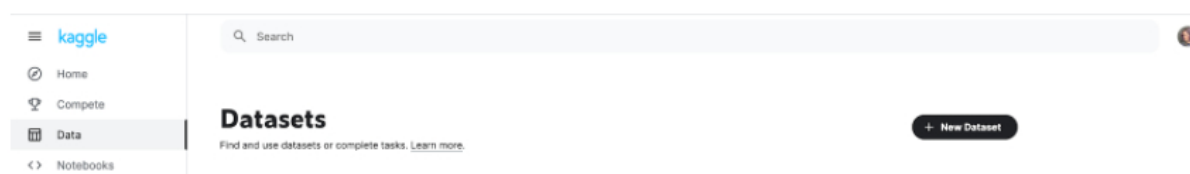
Let's explore the datasets feature!

Find a dataset

1. To start, log in to your Kaggle account.

- **Note:** Kaggle frequently updates its user interface. The latest changes may not be reflected in the screenshots, but the principles in this activity remain the same. Adapting to changes in software updates is an essential skill for data analysts, and we encourage you to practice troubleshooting. You can also reach out to your community of learners on the discussion forum for help.

2. Then, click on the **Data** icon in the **Navigation** bar on the left. This takes you to the **Datasets** home page. From here, you can create a new dataset or search for datasets created by other Kagglers.



Discuss

Courses

More

Recently Viewed

Dive into dplyr (tutorial...

gganimate

gganimate

Writing Hamilton Lyrics...

Street Fighter V - EDA

## Engage With Dataset Tasks

You can now actively engage with datasets with thousands of tasks! Help the community by creating and solving Tasks on datasets!

[See Details](#)

Search 60,966 datasets

Feedback

Filter

Public

Your Datasets

Favorites

Sort by: Hottest

US Election 2020

A. Raphael Fortes

8h

429 KB

10.0

11 Files (CSV)

1 Task

COVID-19 data from John Hopkins University

A. Anthony Goldbloom

12h

2 MB

9.7

10 Files (CSV)

US Election 2020 Tweets

A. March Hui

10d

353 MB

10.0

2 Files (CSV)

1 Task

US Election 2020 - Presidential Debates

A. Heads or Tails

1mo

199 MB

10.0

12 Files (other, CSV)

Election, COVID, and Demographic Data by County

A. Ethan Schwartz

5d

1020 KB

9.7

3 Files (CSV)

3 Tasks

2020 United States presidential election

A. Rado Stokoe

1y

11 MB

4.5

82 Files (CSV, other)

Open Tasks

Mt Cars

9 Submissions - In Mt Cars

Predict U.S. Airbnb Prices

11 Submissions - In U.S. Airbnb Open Data

High value Customers Identifica...

2 Submissions - In UK-High value Customer...

Analysis of Highly Demanded pl...

2 Submissions - In English Premier League T...

Education & Development of BR...

0 Submissions - In BRICS World Bank Indicat...

3. Now, check out a specific dataset. Type **Animal Crossing** in the search bar to find datasets related to the Nintendo video game *Animal Crossing*.

4. There's more than one option, so click on the **Animal Crossing New Horizons Catalog**. This takes you to the landing page for this dataset.

Tour a dataset landing page

kaggle

Home

Compete

Data

Notebooks

Discuss

Courses

More

Recently Viewed

Dive into dplyr (tutorial...

gganimate

gganimate

Writing Hamilton Lyrics...

Street Fighter V - EDA

Search

Animal Crossing New Horizons Catalog

Animal Crossing New Horizons Catalog

A comprehensive inventory of ACNH items, villagers, clothing, fish/bugs etc.

Jessica Li • updated 6 months ago (Version 1)

Data

Tasks (2)

Notebooks (8)

Discussion (2)

Activity

Metadata

Download (577 KB)

New Notebook

Usability 8.2

License CC0 Public Domain

Tags video games, simulations

Description

Context

This dataset comes from this [spreadsheet](#), a comprehensive item Catalog for Animal Crossing New Horizons (ACNH). As described by [Wikipedia](#),

ACNH is a life simulation game released by Nintendo for Nintendo Switch on March 20, 2020. It is the fifth main series title in the Animal Crossing series and, with 5 million digital copies sold, has broken the record for Switch title with most digital units sold in a single month. In New Horizons, the player assumes the role of a customizable character who moves to a deserted island. Taking place in real-time, the player can explore the island in a nonlinear fashion, gathering and crafting items, catching insects and fish, and developing the island into a community of anthropomorphic animals.

Data Explorer

3.47 MB

accessories.csv (50.45 KB)

10 of 22 columns

Detail

Compact

Column

| A Name          | 3"  | A Variation | 3"  | ✓ DRY | 3"             | A Buy       | 3"  | It Sell | 3"   |
|-----------------|-----|-------------|-----|-------|----------------|-------------|-----|---------|------|
| pacifier        | 4%  | Black       | 10% |       | True 1 0%      | 560         | 17% |         |      |
| round shades    | 4%  | Green       | 11% |       | False 221 100% | 660         | 16% |         |      |
| Other (200)     | 93% | Other (164) | 74% |       |                | Other (152) | 68% | 10      | 2020 |
| 3D glasses      |     | White       |     | No    |                | 490         |     | 122     |      |
| 3D glasses      |     | Black       |     | No    |                | 490         |     | 122     |      |
| sandage         |     | Beige       |     | No    |                | 140         |     | 35      |      |
| beak            |     | Yellow      |     | No    |                | 490         |     | 122     |      |
| birthday shades |     | Yellow      |     | No    | N/A            | N/A         |     | 628     |      |
| birthday shades |     | Pink        |     | No    | N/A            | N/A         |     | 628     |      |
| birthday shades |     | Red         |     | No    | N/A            | N/A         |     | 628     |      |

**Header:** The header at the top of the page contains the following information about the dataset:

- Its title

- A brief description of its contents
- The name of its creator
- When it was last updated
- Its current version

**Badge:** In the top-right corner of the header, you'll find three more items:

- A badge in the shape of a circle
- An icon in the shape of a caret symbol ( ^ )
- A number

The badge is related to Kaggle's progression system. If you want, you can read more about it [here](#).

**Upvotes:** Clicking on the caret lets you “upvote” the dataset. The number shows the number of times this dataset has been upvoted by the Kaggle community.

**Tags:** Beneath the header is a bar with six tabs: Data, Tasks, Notebooks, Discussion, Activity, and Metadata. Take a moment to click on each of these tabs and explore their contents. Afterwards, navigate back to the Data tab!

Now, you can move down the page. You'll find a box that contains three terms: Usability, License, and Tags.

**Usability** shows how complete the dataset *webpage* is (and not the dataset itself). Kaggle encourages the community to add information to the dataset webpage to make the dataset itself easier to understand. For example, a brief description or a column header. Hover your cursor over the **Usability score** to discover what the dataset page contains.

**Licenses** govern how a dataset can be used. Click on the **license name** to learn more about that specific license.

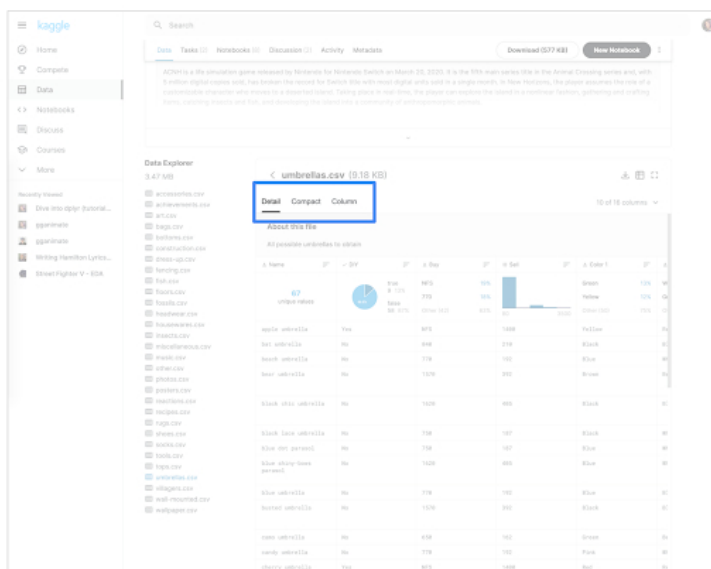
**Tags** refer to different themes or categories. For example, if you click on the **video games** tag, you'll go to a page that shows you everything related to video games on Kaggle. This includes competitions, notebooks, and datasets!

The next box down contains a detailed description of the dataset. Kagglers often include information on where the dataset came from and how the dataset was prepared.

And last—but certainly not least—is the Data Explorer!

## Use the data explorer

The **Data Explorer** menu shows that the *Animal Crossing* dataset contains 30 .csv files. If you click on a file name, the window to the right will display information from that specific file. Try clicking on **umbrellas.csv** to check it out!



## Tour the dataset explorer

---

Notice that the Data Explorer has three viewing options: Detail, Compact, and Column. For now, we'll focus on the Detail tab.

The description at the top of the Detail tab shows that the `umbrellas.csv` file contains data on all the umbrellas in the video game. Let's check out the columns. Each column header has three items:

- A small icon on the left that shows the data type
- The name of the column
- An icon with three bars that lets you sort the data if you click on it

Below each column header is a box that contains a summary of the data. This lets you quickly get an idea of what's in the dataset. For example, the summary for the **Name** column shows there are 67 unique values for the umbrella names. The summary for the **DIY** column shows that 9 of the umbrella recipes are DIY, or "do it yourself." Take a moment to explore the summaries for the other columns.

And that completes our tour! That's a lot of information. Feel free to go back and review.

## Access a dataset

---

After you've explored a dataset, you can link it to a Kaggle notebook or download it to access it for your own use. Linking a dataset to a Kaggle notebook means you create a new notebook from the existing dataset so that it is available for you to use.

### Find your own datasets

---

Now, you'll get a chance to choose your own datasets to work with! Use the following steps to find datasets that interest you:

1. When you're ready, click on the **Data** icon on the left to return to the Datasets landing page.
2. Note that datasets can exist in a variety of formats. If you want to make sure your dataset is in a .csv format, click on the **Filter** button on the right side of the Datasets search bar. Then, choose **CSV** from the menu.
3. Find 2-3 datasets that you're interested in exploring further.
4. Create notebooks from them, download them, or check them out in the Data Explorer. Keep these datasets in mind for your upcoming reflection.

### Link or download a dataset

---

Here are the options to create a notebook or download the dataset:

- Create a Kaggle notebook: To link a dataset to a Kaggle notebook, you click on the **New Notebook** button in the dataset header. This will create a notebook in your Kaggle account that links to the dataset.
- Download the dataset: To download a copy of the dataset to your computer, click on the **Download** button in the dataset header at the top of the page.
- Open the file in Google Sheets: To open a Google Sheets view of the file, click on the **download** icon at the top-right of the Data Explorer. You can then download the file.

## Confirmation and reflection

---

Kaggle's datasets and Data Explorer allow you to do which tasks?

☒ Upload your own datasets

☒ **Correct**

Kaggle's datasets and Data Explorer allow you to search for, access, and upload your own datasets. You can use Kaggle to conduct research, complete data projects, and share your accomplishments with other members of the data science community.

☒ Search for datasets

☒ **Correct**

Kaggle's datasets and Data Explorer allow you to search for, access, and upload your own datasets. You can use Kaggle to conduct research, complete data projects, and share your accomplishments with other members of the data science community.

☐ Create visualizations from datasets

☒ Access datasets

☒ **Correct**

Kaggle's datasets and Data Explorer allow you to search for, access, and upload your own datasets. You can use Kaggle to conduct research, complete data projects, and share your accomplishments with other members of the data science community.

2. So far, you've learned a lot about how to use Kaggle to explore datasets. During this activity, you used this knowledge to find datasets you're interested in. Keep those datasets in mind and in the text box below, write 2-3 sentences (40-60 words) in response to each of the following questions:

1 / 1 point

- What type(s) of data are in this dataset? Is this dataset biased or unbiased? How do you know?
- Based on what you've explored so far, how might you use Kaggle's Datasets feature to help develop your data analysis skills?

What type(s) of data are in this dataset? Is this dataset biased or unbiased? How do you know?

There are data from the Animal Crossing New Horizons Catalog. This data is probably unbiased my conclusion is based on the dataset's trustworthiness.

Based on what you've explored so far, how might you use Kaggle's Datasets feature to help develop your data analysis skills?

It will help me by offering a platform with an active community.

☒ **Correct**

Congratulations on completing this hands-on activity! A strong response would include that Kaggle gives you access to tons of public datasets that you can use to practice data analysis and create your own projects. Beyond that, consider the following:

Online platforms like Kaggle allow you to search for, view, explore, upload, and work with datasets from a variety of sources and perspectives. Understanding how Kaggle works and how to use it will help you develop your skillset and grow as a data analyst.