



Photo credit: [SeventyFour](#)

What is a database?

Databases store and organize data electronically. The data is housed in objects called tables, with data organized into rows and columns (or records and fields).

A relational database defines relationships between tables of data inside a database.

Although they can look similar to a spreadsheet, databases can store more data, more securely, and allow multiple users to access the same database.

Our data

This project will use data from the [Austin Animal Shelter](#). In this case, the data is saved to this DataLab workbook as a CSV file, but you'll query it just like you would a database!

Here is copy of the data description from the Animal Shelter website and what you'd expect to find in the data:

Column Name	Description	Data Type
animal_id	Identifier	Text
name	Animal name, if provided	Text
datetime	Date of outcome	Floating Timestamp
monthyear	Month and year of outcome	Floating Timestamp
date_of_birth	Animal's date of birth	Floating Timestamp
outcome_type	Details of the outcome	Text
outcome_subtype	Additional outcome details	Text
animal_type	Type of animal	Text
sex_upon_outcome	Details on whether the animal was neutered or intact	Text
age_upon_outcome	Age of animal at time of outcome	Text
breed	Animal breed	Text
color	Animal color	Text

Step 1: Explore the data

Before you dig into the data, it's always a good idea to explore it to see if it is as expected. Does the data match what was outlined above? Try querying the data to view a snapshot of the first 10 rows.

◻ DataFrames and CSVs DataFrame as `animal_shelter`

```
SELECT *
FROM 'animal_shelter.csv'
LIMIT 10;
```

...	↑↓	a	...	↑↓	n	...	↑↓	datetime	...	↑↓	nr	...	↑↓	date_of_birth	...	↑↓	outc...	...	↑↓	outcome_...	...
0		A882831	*	Hamilton		*Hamilton		2023-07-01T18:12:00.000			Jul 2023			2023-03-25T00:00:00.000			Adoption			null	
1		A794011	Chunk			Chunk		2019-05-08T18:20:00.000			May 2019			2017-05-02T00:00:00.000			Rto-Adopt			null	
2		A776359	Gizmo			Gizmo		2018-07-18T16:02:00.000			Jul 2018			2017-07-12T00:00:00.000			Adoption			null	
3		A821648	null			null		2020-08-16T11:38:00.000			Aug 2020			2019-08-16T00:00:00.000			Euthanasia			null	
4		A720371	Moose			Moose		2016-02-13T17:59:00.000			Feb 2016			2015-10-08T00:00:00.000			Adoption			null	
5		A674754	null			null		2014-03-18T11:47:00.000			Mar 2014			2014-03-12T00:00:00.000			Transfer			Partner	
6		A659412	Princess			Princess		2020-10-05T14:37:00.000			Oct 2020			2013-03-24T00:00:00.000			Adoption			null	
7		A814515	Quentin			Quentin		2020-05-06T07:59:00.000			May 2020			2018-03-01T00:00:00.000			Adoption			Foster	
8		A902098	null			null		2024-07-18T12:07:00.000			Jul 2024			2024-04-10T00:00:00.000			Adoption			Foster	
9		A860161	*Lalo			*Lalo		2022-07-19T15:53:00.000			Jul 2022			2022-05-04T00:00:00.000			Adoption			null	

Rows: 10

It's also helpful to understand how big the dataset is. How many rows (or records) are there in this data? Query the file to find out.

◻ DataFrames and CSVs DataFrame as `animal_shelter_2`

```
SELECT COUNT(*) AS total_rows
FROM 'animal_shelter.csv';
```

index	...	↑↓	total_rows
			0

Rows: 1

Your turn! Can you think of another way to count the rows (or records)?

◻ DataFrames and CSVs DataFrame as `animal_s1`

```
SELECT COUNT(animal_type) AS count_type
FROM 'animal_shelter.csv';
```

index	...	↑↓	count_type
			0

Rows: 1

Step 2: Get deeper and answer questions about the data

Let's explore the data further. You're probably curious to know what animals came through this shelter so let's start there. How many different animal types has this shelter taken in?

◻ DataFrames and CSVs DataFrame as `animal_u`

```
SELECT COUNT(DISTINCT animal_type) AS unique_type_count
FROM 'animal_shelter.csv';
```

index	...	↑↓	unique_type_count
			0

Rows: 1

Now you know how many different animal types there were, can you also find out some more information? What were those animals and how many were there of each animal type?

◻ DataFrames and CSVs DataFrame as animal

```
SELECT animal_type, COUNT(animal_type) AS count_type
FROM 'animal_shelter.csv'
GROUP BY animal_type;
```

index	animal_type	count_type
0	Dog	
1	Bird	
2	Other	
3	Cat	
4	Livestock	

Rows: 5

You may have noticed in our first investigation of the data that some animal names were listed as null. This means they didn't have a name or their name is missing from the records for another reason.

Your turn! How many animals don't have a name?

◻ DataFrames and CSVs DataFrame as

```
SELECT COUNT(*) AS total_rows
FROM 'animal_shelter.csv'
WHERE name IS NULL;
```

total_rows
48655

Rows: 1

Were the animals without a name all a specific animal type?

◻ DataFrames and CSVs DataFrame as

```
SELECT DISTINCT animal_type
FROM 'animal_shelter.csv'
WHERE name IS NULL
```

animal_type
Dog
Bird
Other
Cat
Livestock

Rows: 5

◻ DataFrames and CSVs DataFrame as

```
SELECT DISTINCT animal_type, COUNT(*) AS count_animals
FROM 'animal_shelter.csv'
WHERE name IS NULL
GROUP BY animal_type;
```

animal_type	count_animals
Other	7210
Cat	26270
Dog	14510
Bird	636
Livestock	29

Rows: 5

Livestock and other seem like broad categories. Do we have any more details on those without a name?

DataFrames and CSVs DataFrame as

```
SELECT animal_type, breed
FROM 'animal_shelter.csv'
WHERE name IS NULL
    AND animal_type IN ('Other', 'Livestock')
GROUP BY animal_type, breed
ORDER BY animal_type;
```

...	↑↓ ani...	...	↑↓	breed	...	↑↓
0	Livestock			Pig		
1	Livestock			Sheep Mix		
2	Livestock			Potbelly Pig Mix		
3	Livestock			Pygmy		
4	Livestock			Goat/Pygmy		
5	Livestock			Goat Mix		
6	Livestock			Goat		
7	Livestock			Pig Mix		
8	Livestock			Potbelly Pig		
9	Livestock			Miniature		
10	Livestock			Emu		
11	Other			Opossum		
12	Other			Chinchilla		
13	Other			Fox Mix		
14	Other			Cold Water		
15	Other			Hamster Mix		
16	Other			Mouse Mix		

Rows: 102

Step 3: It's raining cats and dogs

It's probably safe to say that cats and dogs are the most popular pets. But, what is the most common cat breed this shelter saw?

DataFrames and CSVs DataFrame as

```
SELECT breed, COUNT(breed) AS count_breed
FROM 'animal_shelter.csv'
WHERE animal_type IN ('Cat')
GROUP BY breed
ORDER BY count_breed DESC
LIMIT 1;
```

...	↑↓ breed	...	↑↓	co... co...	...	↑↓
0	Domestic Shorthair Mix			34006		

Rows: 1

What were the most common dog names in 2023?

DataFrames and CSVs DataFrame as

```
SELECT name, COUNT(*) AS name_count
FROM 'animal_shelter.csv'
WHERE animal_type = 'Dog'
    AND monthyear LIKE '%2023%'
    AND name IS NOT NULL
GROUP BY name
ORDER BY name_count DESC
LIMIT 5;
```

...	↑↓	... n...	... ↑↓	...	↑↓
0	Luna			38	
1	Max			24	
2	Bella			21	
3	Milo			17	
4	Daisy			17	

Rows: 5

How many animals were adopted in 2023?

DataFrames and CSVs DataFrame as

```
SELECT COUNT(*) AS name_count
FROM 'animal_shelter.csv'
WHERE monthyear LIKE '%2023%'
    AND outcome_type = 'Adoption';
```

...	↑↓	n...	... ↑↓
0		6634	

Rows: 1

Your turn! What was the most adopted animal type and breed in 2023?

DataFrames and CSVs DataFrame as

```
SELECT animal_type, breed, COUNT(*) AS adopted
FROM 'animal_shelter.csv'
WHERE monthyear LIKE '%2023%'
    AND outcome_type = 'Adoption'
GROUP BY animal_type, breed
ORDER BY adopted DESC
LIMIT 10;
```

...	↑↓	ani... ↑↓	breed	...	↑↓	... ↑↓
0	Cat			Domestic Shorthair		2709	
1	Cat			Domestic Shorthair Mix		280	
2	Dog			Pit Bull		271	
3	Dog			Labrador Retriever Mix		262	
4	Cat			Domestic Medium Hair		215	
5	Dog			Pit Bull Mix		214	
6	Dog			German Shepherd Mix		172	
7	Dog			German Shepherd		129	
8	Dog			Labrador Retriever		116	
9	Cat			Domestic Longhair		83	

Rows: 10