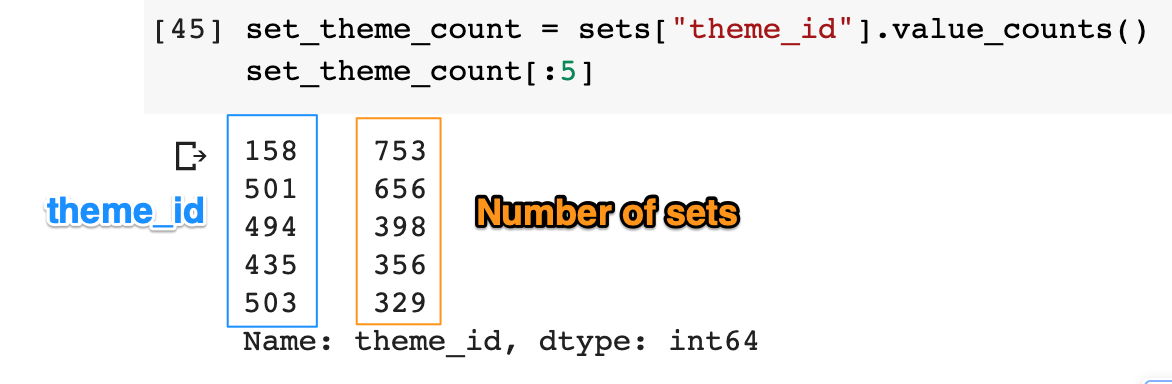
**Relational Database Schemas: Primary and Foreign Keys**

LEGO has licensed many hit franchises from Harry Potter to Marvel Super Heros to many others. But which theme has the largest number of individual sets? Is it one of LEGO's own themes like Ninjago or Technic or is it a third party theme? Let's analyse LEGO's product lines in more detail!

**Number of Sets per LEGO Theme**

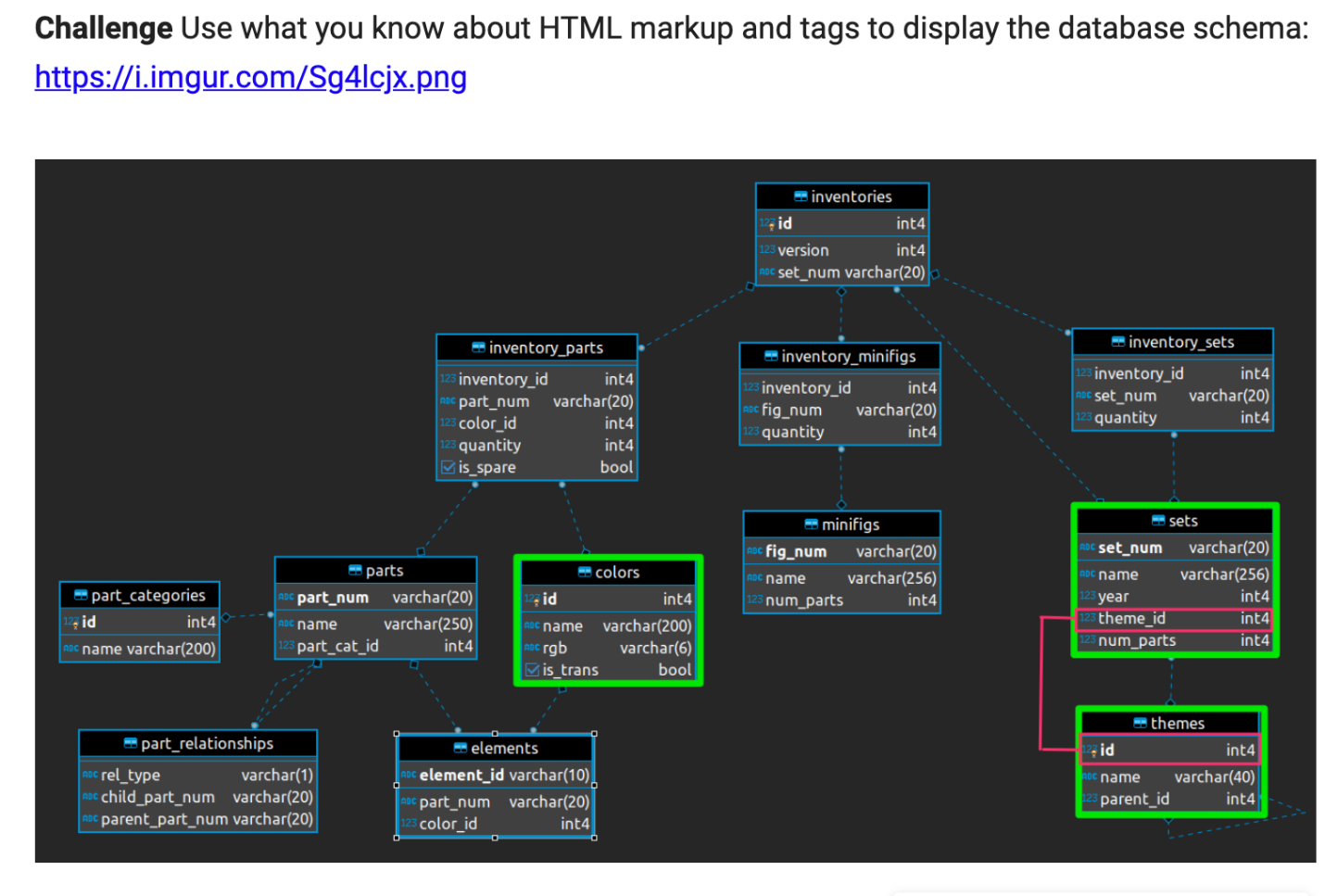
To count the number of sets per Theme we can use the .value\_counts() method on our theme\_id column. But there's one problem:



We have no idea what our themes are actually called! 🤨 Ok, we can see that the theme with id 158 is the largest theme containing 753 individual sets, but what's that theme called? This is not very helpful. We need to find the names of the themes based on the theme\_id from the themes.csv file.

**Mini-Challenge**

Display the database schema (link: https://i.imgur.com/Sg4lcjx.png) inside the Notebook. You should see something like this:



**Solution: Embedding images in a Markdown Cell:**

To display an image in a Text (aka Markdown) cell, all you need to do is use the HTML <img> tag.

1. <img src="https://i.imgur.com/Sg4lcjx.png">

**Working with a Relational Database**

What is a database schema? A schema is just how the database is organised. Many relational databases, such as our LEGO data, is split into individual tables. We have separate tables for the colours, the sets and the thems. With a relational database, the tables are linked to each other through their keys.

**Understand the theme.csv file**

The themes.csv file has the actual theme names. How is this table linked to the others tables? Well, the sets .csv has theme\_ids which match the id column in the themes.csv.

This means that the theme\_id is the **foreign key** inside the sets.csv. Many different sets can be part of the same theme. But *inside* the themes.csv, each theme\_id, which is just called id is unique. This uniqueness makes the id column the **primary key** inside the themes.csv. To see this in action, explore the themes.csv.

Challenge

How is the themes.csv structured?

Search for the name 'Star Wars'. How many ids correspond to the 'Star Wars' name in the themes.csv?

Use the ids you just found and look for the corresponding sets in the sets.csv (Hint: you'll need to look for matches in the theme\_id column).

.

..

...

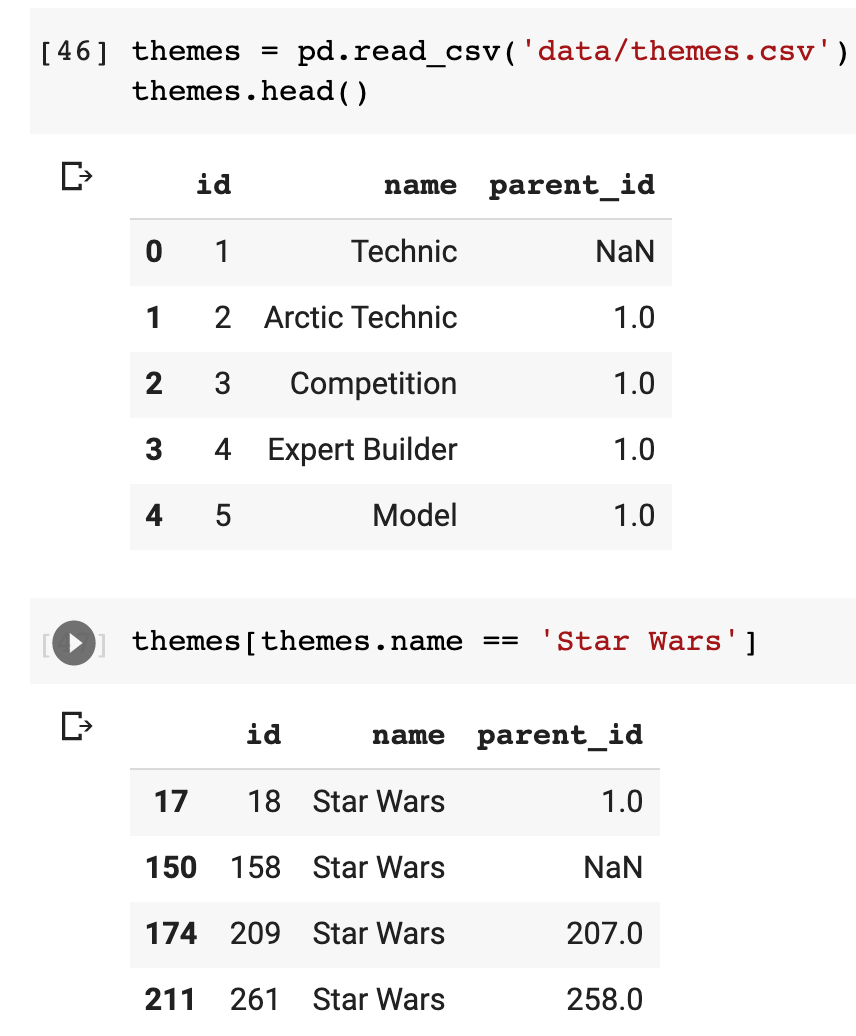
...

..

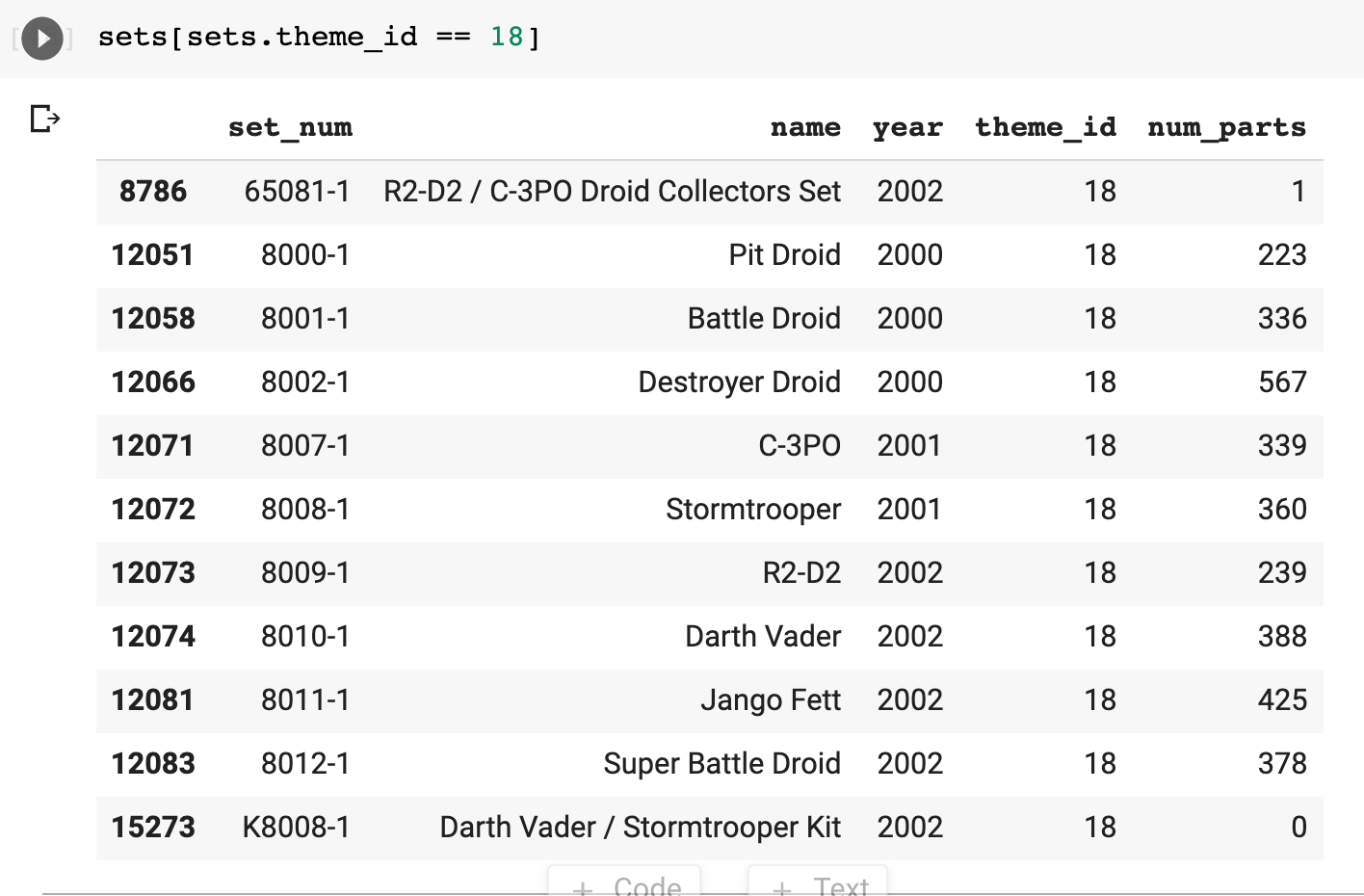
.

**Solution: Exploring the themes.csv**

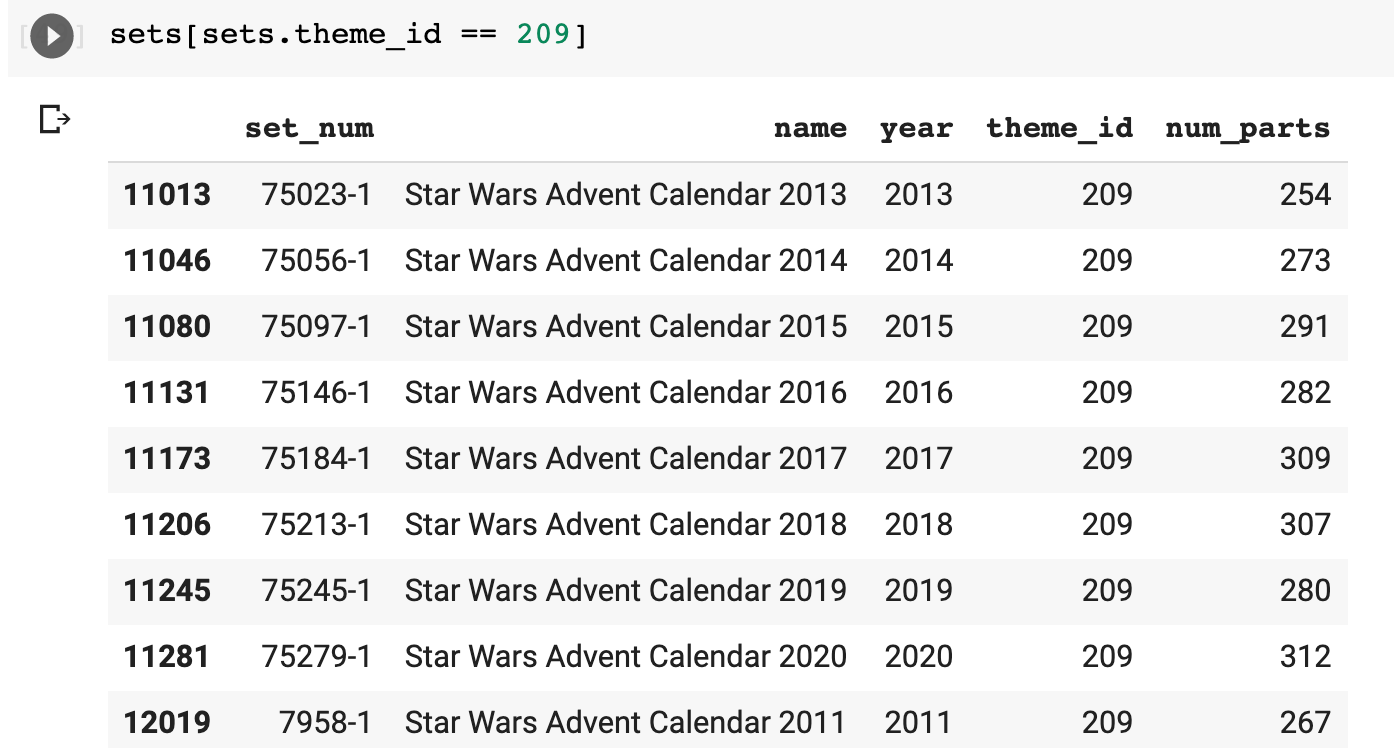
Looking at the first 5 rows, we see the column names. Each value in the id column is unique (this is the primary key for the themes table). The theme names are not unique. If we search for the name "Star Wars", we see that 4 different ids correspond to that name.



Why would Star Wars have so many different themes? We can check which products corresponded to those themes in the sets.csv:



Star Wars is a really long-running franchise. Theme number 18 was running from 2000 to 2002 and seems to be comprised of several of the show's characters. What about, say theme 209?



Here we see that all of the Star Wars Advent Calendars share the same theme\_id. That makes sense.

