

## Using SQL to clean data

### Learn basic SQL queries

#### Transforming data

- Reading: Overview: Upload the store transactions dataset to BigQuery 10 mins
- Video: Advanced data cleaning: Averages, part 1 6 min
- Video: Advanced data cleaning: Functions, part 1 8 min
- Discussion Prompt: Debugging SQL code 10 mins
- Updated Playlist: Data cleaning with SQL 38 mins
- Practice Quiz: Self-reflection: Challenges with SQL 1 question

#### Weekly challenge 3

## Optional: Upload the store transactions dataset to BigQuery

In the next video, the instructor uses a specific dataset. The instructions in this reading are provided for you to upload the same dataset to your BigQuery console so you can follow along.

You must have a BigQuery account to follow along. If you have happened around courses, [Link to BigQuery](#). It's in the **Prepare Data for Exploration** course covers how to set up a BigQuery account.

### Prepare for the next video

- First, download the CSV file from the attachment below.

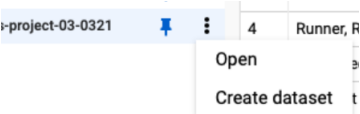
Lauren's Furniture Store Transaction Table  
CSV File

- Next, complete the steps below in your BigQuery console to upload the Store Transaction dataset.

**Note:** These steps will be different from what you performed before. In previous instances, you selected the Auto detect check box to allow BigQuery to auto-detect the schema. This time, you will choose to create the schema by editing it as text. This method can be used when BigQuery doesn't automatically set the desired type for a particular field. In this case, you will specify STRING instead of FLOAT as the type for the purchase\_price field.

**Step 8:** Open your BigQuery console and click on the project you want to upload the data to. If you already created a **customer\_data** dataset for your project, jump to step 5; otherwise, continue with step 2.

**Step 9:** In the Explorer on the left, click the Actions icon (three vertical dots) next to your project name and select **Create dataset**.



**Step 9:** Enter **customer\_data** for the Dataset ID.

### Create dataset

**Dataset ID \***  
customer\_data  
Letters, numbers, and underscores allowed

**Data location**  
Default

### Default table expiration

☐ Enable table expiration  
Default maximum table age: Days

### Encryption

- ☒ Google-managed encryption key  
No configuration required
- ☐ Customer-managed encryption key (CMEK)  
Manage via Google Cloud Key Management Service

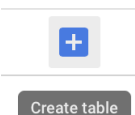
CREATE DATASET CANCEL

**Step 4:** Click **CREATE DATASET** (blue button) to add the dataset to your project.

**Step 5:** In the Explorer, click to expand your project, and then click the **customer\_data** dataset.

**Step 6:** Click the Actions icon (three vertical dots) next to customer\_data and select **Open**.

**Step 7:** Click the blue + icon at the top right to open the Create table window.



**Step 8:** Under **Source**, for the Create table from selection, choose where the data will be coming from.

- Select **Upload**.
- Click **Browse** to select the Store Transaction Table CSV file you downloaded.
- Choose **CSV** from the file format drop-down.

### Create table

**Source**  
Create table from: Upload  
Select file: Lauren's Furniture Store Transaction Table.csv  
File format: CSV

**Destination**  
Search for a project: Enter a project name  
Project name: Lauren's project 03-0321  
Dataset name: customer\_data  
Table type: Table

**Table name**  
customer\_purchase

**Schema**  
Auto detect  
Schema and input parameters

☐ Edit as text

**Step 9:** For Table name, enter **customer\_purchase** if you plan to follow along with the video.

**Step 10:** For Schema, click the toggle switch for **Edit as text**. This opens up a box for the text.

**Step 11:** Copy and paste the following text into the box. Be sure to include the opening and closing brackets. They are required.

```
{
  "description": "date",
  "mode": "NULLABLE",
  "name": "date",
  "type": "DATETIME"
},
{
  "description": "transaction id",
  "mode": "NULLABLE",
  "name": "transaction_id",
  "type": "INTEGER"
},
{
  "description": "customer id",
  "mode": "NULLABLE",
  "name": "customer_id",
  "type": "INTEGER"
},
{
  "description": "product name",
  "mode": "NULLABLE",
  "name": "product",
  "type": "STRING"
},
{
  "description": "product code",
  "mode": "NULLABLE",
  "name": "product_code",
  "type": "STRING"
},
{
  "description": "product color",
  "mode": "NULLABLE",
  "name": "product_color",
  "type": "STRING"
},
{
  "description": "product price",
  "mode": "NULLABLE",
  "name": "product_price",
  "type": "FLOAT"
},
{
  "description": "quantity purchased",
  "mode": "NULLABLE",
  "name": "purchase_size",
  "type": "INTEGER"
},
{
  "description": "purchase price",
  "mode": "NULLABLE",
  "name": "purchase_price",
  "type": "STRING"
},
{
  "description": "revenue",
  "mode": "NULLABLE",
  "name": "revenue",
  "type": "FLOAT"
}
}
```

**Step 12:** Scroll down and expand the **Advanced options** section.

**Step 13:** For the **Header rows to skip** field, enter 1.

**Advanced options**  
Write preference: Write all empty  
Number of errors allowed: 0  
Field delimiter: Comma  
Header rows to skip: 1  
Quoted newlines: Allow quoted newlines  
Jagged rows: Allow jagged rows  
Encryption: Google-managed key (No configuration required)  
Customer-managed key (Manage via Google Cloud Key Management Service)

**Step 14:** Click **Create table** (blue button). You will now see the **customer\_purchase** table under your **customer\_data** dataset in your project.

**Step 15:** Click the **customer\_purchase** table and in the **Schema tab**, confirm that the schema matches the schema shown below.

SCHEMA

DETAILS

PREVIEW

Table schema

Filter

Enter property name or value

Field name	Type	Mode	Policy Type	Description
date	DATETIME	NULLABLE		date
transaction_id	INTEGER	NULLABLE		transaction id
customer_id	INTEGER	NULLABLE		customer id
product	STRING	NULLABLE		product name
product_code	STRING	NULLABLE		product code
product_color	STRING	NULLABLE		product color
product_price	FLOAT	NULLABLE		product price
purchase_size	INTEGER	NULLABLE		quantity purchased
purchase_price	STRING	NULLABLE		purchase price
revenue	FLOAT	NULLABLE		revenue

**Step 16:** Click the **Preview tab** and confirm that your data matches the data shown below.

ROW	date	transaction_id	customer_id	product	product_code	product_color	product_price	purchase_size	purchase_price	revenue
1	2020-03-07T00:00:00	9890	1880	toy	00400000	black	10.99	2	21.98	21.98
2	2020-03-07T00:00:00	13375	2940	toy	00400000	black	10.99	2	21.98	21.98
3	2020-03-07T00:00:00	9890	1730	toy	00400000	black	10.99	1	10.99	10.99
4	2020-03-07T00:00:00	48970	3900	toy	00400000	black	10.99	1	10.99	10.99
5	2020-03-07T00:00:00	48970	3940	toy	00400000	black	10.99	2	21.98	21.98
6	2020-03-07T00:00:00	48970	3940	toy	00400000	black	40.99	2	81.98	81.98
7	2020-03-07T00:00:00	10280	2010	bed	00400000	black	10.99	1	10.99	10.99
8	2020-03-07T00:00:00	9440	1830	sofa	00300000	gray	1000.00	1	1000.00	1000.00
9	2020-03-07T00:00:00	20400	4040	sofa	00300000	gray	1000.00	1	1000.00	1000.00
10	2020-03-07T00:00:00	49400	3940	sofa	00300000	gray	1000.00	1	1000.00	1000.00
11	2020-03-07T00:00:00	49400	3910	sofa	00300000	gray	1000.00	1	1000.00	1000.00
12	2020-03-07T00:00:00	49400	3910	sofa	00300000	gray	1000.00	1	1000.00	1000.00
13	2020-03-07T00:00:00	49400	3910	sofa	00300000	gray	1000.00	1	1000.00	1000.00
14	2020-03-07T00:00:00	1470	3310	sofa	00300000	gray	1000.00	1	1000.00	1000.00
15	2020-03-07T00:00:00	20400	4040	sofa	00300000	gray	1000.00	1	1000.00	1000.00
16	2020-03-07T00:00:00	19100	3810	sofa	00300000	gray	1000.00	1	1000.00	1000.00
17	2020-03-07T00:00:00	49400	3910	toy	00400000	black	10.99	2	21.98	21.98
18	2020-03-07T00:00:00	21470	4330	toy	00400000	gray	80.00	2	160.00	160.00
19	2020-03-07T00:00:00	21470	4330	toy	00400000	gray	80.00	1	80.00	80.00
20	2020-03-07T00:00:00	43400	3840	sofa	00400000	gray	1000.00	1	1000.00	1000.00
21	2020-03-07T00:00:00	43400	3840	sofa	00400000	gray	1000.00	1	1000.00	1000.00
22	2020-03-07T00:00:00	43400	3840	sofa	00400000	gray	1000.00	1	1000.00	1000.00
23	2020-03-07T00:00:00	43400	3840	sofa	00400000	gray	1000.00	1	1000.00	1000.00
24	2020-03-07T00:00:00	14400	2910	sofa	00400000	gray	1000.00	1	1000.00	1000.00
25	2020-03-07T00:00:00	14400	2910	sofa	00400000	gray	1000.00	1	1000.00	1000.00
26	2020-03-07T00:00:00	14400	2910	sofa	00400000	gray	1000.00	1	1000.00	1000.00
27	2020-03-07T00:00:00	21400	4270	sofa	00400000	gray	200.00	1	200.00	200.00
28	2020-03-07T00:00:00	21400	4270	sofa	00400000	gray	200.00	1	200.00	200.00
29	2020-03-07T00:00:00	21400	4270	sofa	00400000	gray	200.00	1	200.00	200.00

Congratulations, you are now ready to follow along with the video!

### Mark as completed

Like dislike Report an issue