

Using SQL to clean data

Learn basic SQL queries

Transforming data

Weekly challenge 3

Reading: Glossary: Terms and definitions

5 min

Quiz: Weekly challenge 3

9 questions

Congratulations! You passed!

Grade received

Quiz • 45 min

Review Learning Objectives

1. A data analyst is analyzing medical data for a health insurance company. The dataset contains billions of rows of data. Which of the following tools will handle the data most efficiently?

1 / 1 point

Submit your assignment

Due Jul 2, 11:59 PM +00 Attempts 3 every 24 hours

CSV

SQL

A word processor

A spreadsheet

To Pass 80% or higher

Correct

2. Your manager asks you with analyzing a database and visually inspecting the data. Upon initial inspection you realize that this is a small dataset. What tool should you use to analyze the data?

1 / 1 point

Word processor

CSV

Spreadsheet

SQL

Correct

3. A data analyst creates many new tables in their company's database. When the project is complete, the analyst wants to remove the tables so they don't clutter the database. What SQL commands can they use to delete the tables?

1 / 1 point

DROP TABLE IF EXISTS

CREATE TABLE IF NOT EXISTS

UPDATE

INSERT INTO

Correct

4. You are working with a database table that contains invoice data. The table includes columns for invoice\_id and billing\_state. You want to remove duplicate entries for billing\_state and sort the results by invoice\_id.

1 / 1 point

You write the SQL query below. Add a DISTINCT clause that will remove duplicate entries from the billing\_state column.

NOTE: The three dots (...) indicate where to add the clause.

```
1 SELECT DISTINCT billing_state
2 FROM Invoice
3 ORDER BY invoice_id;
```

Run

Reset

billing\_state
None |
AZ |
CA |
Dublin |
IL |
IN |
MI |
NV |
NY |
NSW |
PA |
TX |
WI |
WV |
-----|
(Output limit exceeded, 25 of 26 total rows shown)

What billing state appears in row 17 of your query result?

NOTE: The query index starts at 1 not 0.

CA

NV

AZ

WI

Correct

The clause DISTINCT billing\_state will remove duplicate entries from the billing\_state column. The complete query is SELECT DISTINCT billing\_state FROM Invoice ORDER BY invoice\_id.

The DISTINCT clause removes duplicate entries from your query result. The billing state AZ appears in row 17 of your query result.

5. You are working with a database table that contains customer data. The table includes columns about customer location such as city, state, country, and postal\_code. You want to find what state names are greater than 3 characters.

1 / 1 point

You write the SQL query below. Add a LENGTH function that will return any state names that are greater than 3 characters long.

NOTE: The three dots (...) indicate where to add the clause.

```
1 SELECT *
2 FROM Customer
3 WHERE LENGTH(state) > 3;
```

Run

Reset

customer\_id	first\_name	last\_name	company	address	city	state	country	postal\_code	phone	fax	email	support\_no\_id
1 | 46 | Hugh | O'Reilly | None | 3 Chatham Street | Dublin | Dublin | Ireland | None | +353 01 6792424 | None | hughoreilly@people.ie | 3 |

What state is in row 1 of your query result? (Hint: you will have to scroll to the right with your mouse or track pad to locate the indicated column.)

NOTE: The query index starts at 1 not 0.

Ireland

Chile

India

Dublin

Correct

The function LENGTH(state) > 3 will return any state names that are greater than 2 characters long. The complete query is SELECT \* FROM Customer WHERE LENGTH(state) > 3. The LENGTH function counts the number of characters a string contains. The country Ireland is in row 1 of your query result.

6. In SQL databases, what data type is the value 78.99 an example of?

1 / 1 point

Boolean

Integer

Float

String

Correct

7. In SQL databases, what function can be used to convert data from one datatype to another?

1 / 1 point

SUBSTR

CAST

LENGTH

TRIM

Correct

8. After joining multiple tables you find your data contains a significant amount of null values. What function can you use to return only the non-null values in a list?

1 / 1 point

TRIM

CONCAT

COALESCE

CAST

Correct

9. You are working with a database table that contains employee data. The table includes columns about employee location such as city, state, country, and postal\_code. You use the SUBSTR function to retrieve the first 3 characters of each last\_name, and use the AS command to store the result in a new column called new\_last\_name.

1 / 1 point

You write the SQL query below. Add a statement to your SQL query that will retrieve the first 3 characters of each last\_name and store the result in a new column as new\_last\_name.

NOTE: The three dots (...) indicate where to add the statement.

NOTE: SUBSTR takes in three arguments being column, starting\_index, ending\_index

```
1 SELECT
2   employee_id,
3   SUBSTR(last_name, 1, 3) AS new_last_name
4 FROM
5   employen
6 ORDER BY
7   postal_code;
```

Run

Reset

employee\_id	new\_last\_name
8 | Cal |
7 | Edo |
2 | Edw |
4 | Pav |
3 | Pia |
6 | Mel |
5 | Jon |
1 | Ada |

What employee ID number is in row 8 of your query result?

NOTE: The query index starts at 1 not 0.

7

8

1

3

Incorrect

Please review the video on string variables in SQL 12.