

# 4,000 customers. Or 1,900. Same table.

D8A.ACADEMY/RESOURCES/COUNT-STAR-  
VS-DISTINCT

## THREE COUNTS

|                     |                            |
|---------------------|----------------------------|
| COUNT(*)            | rows, NULLs included       |
| COUNT(col)          | rows where col is not NULL |
| COUNT(DISTINCT col) | different values of col    |

## REMEMBER

|                       |                              |
|-----------------------|------------------------------|
| COUNT(*) - COUNT(col) | the number of NULLs in col   |
| COUNT(1)              | same as COUNT(*), not faster |

## THE REPORT

|                      |                                 |
|----------------------|---------------------------------|
| COUNT(*) = 4,000     | ← order lines, not customers    |
| DISTINCT customer_id | 1,900 people. the real number   |
| the right column     | ← the one that means one person |

## WORTH MEMORIZING

### 01 Same table, three numbers

One query, three counts, three different questions answered.

```
SELECT
  COUNT(*)           AS order_lines,
  COUNT(coupon_code) AS lines_with_coupon,
  COUNT(DISTINCT customer_id) AS customers
FROM campaign_orders;

-- 4000 | 2600 | 1900
```

→ The 4,000 is order lines. The 2,600 is lines where the coupon column is not NULL. The 1,900 is people. The report said customers and showed the first number.

### 02 Trap: DISTINCT on the wrong column

DISTINCT counts different values of whatever you give it. Give it the column that means one person.

| DISTINCT ON | YOU GET | BECAUSE                      |
|-------------|---------|------------------------------|
| order_id    | 3,100   | one customer, several orders |
| email       | 1,950   | one person, two spellings    |
| customer_id | 1,900   | one person, one id           |

→ Say what one row means before you count. One line, one order, one customer: three COUNTs, three numbers, and only one of them answers the question that was asked.