

```

app_db=# -- Gesamtzeilen
SELECT COUNT(*) AS gesamt
FROM retail_raw_v2;

gesamt
-----
541909
(1 row)

app_db=# SELECT
    COUNT(DISTINCT invoiceno) AS distinct_invoices,
    COUNT(DISTINCT customerid) AS distinct_customers,
    COUNT(DISTINCT stockcode) AS distinct_products
FROM retail_raw_v2;
distinct_invoices | distinct_customers | distinct_products
-----+-----+-----
      25900 |          4372 |          4070
(1 row)

app_db=#

```

```

app_db=# SELECT
    COUNT(*) AS gesamt,
    SUM(CASE WHEN invoiceno IS NULL THEN 1 ELSE 0 END) AS invoiceno_null,
    SUM(CASE WHEN stockcode IS NULL THEN 1 ELSE 0 END) AS stockcode_null,
    SUM(CASE WHEN description IS NULL THEN 1 ELSE 0 END) AS description_null,
    SUM(CASE WHEN quantity IS NULL THEN 1 ELSE 0 END) AS quantity_null,
    SUM(CASE WHEN invoicedate IS NULL THEN 1 ELSE 0 END) AS invoicedate_null,
    SUM(CASE WHEN unitprice IS NULL THEN 1 ELSE 0 END) AS unitprice_null,
    SUM(CASE WHEN customerid IS NULL THEN 1 ELSE 0 END) AS customerid_null,
    SUM(CASE WHEN country IS NULL THEN 1 ELSE 0 END) AS country_null
FROM retail_raw_v2;
-[ RECORD 1 ]-----+
gesamt      | 541909
invoiceno_null | 0
stockcode_null | 0
description_null | 1454
quantity_null | 0
invoicedate_null | 0
unitprice_null | 0
customerid_null | 135080
country_null | 0

```

gesamt	invoiceno_null	stockcode_null	description_null	quantity_null	invoicedate_null	unitprice_null	customerid_null	country_null
541909	0	0	1454	0	0	0	0	0

```
app_db=# SELECT
    COUNT(*) AS alle,
    SUM(CASE WHEN quantity < 0 THEN 1 ELSE 0 END) AS negative_mengen
FROM retail_raw_v2;
-[ RECORD 1 ]---+-----
alle           | 541909
negative_mengen | 10624
```

```
app_db=# SELECT
    COUNT(*)                               AS alle,
    SUM(CASE WHEN unitprice < 0 THEN 1 END) AS negative_preise,
    SUM(CASE WHEN unitprice = 0 THEN 1 END) AS zero_preise
FROM retail_raw_v2;
-[ RECORD 1 ]---+-----
alle           | 541909
negative_preise | 2
zero_preise    | 2519
```

```
app_db=# -- Gibt es noch leere Strings? (sollte 0 sein, weil INTEGER)
SELECT COUNT(*)
FROM retail_raw_v2
WHERE customerid::text = '';
-[ RECORD 1 ]
count | 0
```

```
app_db=# -- Anzahl in customers
SELECT COUNT(*) AS dim_customers
FROM customers;
-[ RECORD 1 ]---+-----
dim_customers | 4372

app_db=# -- Anzahl distinct Kunden in Raw
SELECT COUNT(DISTINCT customerid) AS raw_customers
FROM retail_raw_v2
WHERE customerid IS NOT NULL;
-[ RECORD 1 ]---+-----
raw_customers | 4372
```

```
app_db=# -- Gibt es Verkäufe mit CustomerID, die NICHT in customers stehen?
SELECT COUNT(*) AS fehlende_kunden
FROM retail_raw_v2 r
LEFT JOIN customers c USING (customerid)
WHERE r.customerid IS NOT NULL
    AND c.customerid IS NULL;
-[ RECORD 1 ]---+-
fehlende_kunden | 0
```

```

app_db=# -- Anzahl in products
SELECT COUNT(*) AS dim_products
FROM products;
-[ RECORD 1 ]+-----
dim_products | 4070

app_db=# -- Anzahl distinct Stockcodes im Raw
SELECT COUNT(DISTINCT stockcode) AS raw_products
FROM retail_raw_v2
WHERE stockcode IS NOT NULL;
-[ RECORD 1 ]+-----
raw_products | 4070

```

```

app_db=# -- Hat ein Stockcode mehrere verschiedene Beschreibungen?
SELECT stockcode,
       COUNT(DISTINCT description) AS verschiedene_beschreibungen
FROM retail_raw_v2
WHERE stockcode IS NOT NULL
GROUP BY stockcode
HAVING COUNT(DISTINCT description) > 1
LIMIT 20;
stockcode | verschiedene_beschreibungen
-----+-----
10080      |          2
10133      |          2
15058A     |          2
15058C     |          2
16008      |          2
16045      |          2
16156L     |          2
16162M     |          2
16168M     |          2
16169E     |          2
16207A     |          2
17012B     |          2
17012C     |          2
17012F     |          2
17091A     |          2
17107D     |          3
17109D     |          2
17136A     |          2
20622      |          2
20652      |          2
(20 rows)

```

```
app_db=# -- Gibt es Stockcodes im Raw, die NICHT in products gelandet sind?
SELECT COUNT(*) AS fehlende_products
FROM retail_raw_v2 r
LEFT JOIN products p USING (stockcode)
WHERE r.stockcode IS NOT NULL
    AND p.stockcode IS NULL;
fehlende_products
-----
          0
(1 row)
```

```
app_db=# -- Anzahl in invoices
SELECT COUNT(*) AS dim_invoices
FROM invoices;
dim_invoices
-----
      22190
(1 row)

app_db=# -- Anzahl distinct Rechnungen mit CustomerID in Raw
SELECT COUNT(DISTINCT invoiceno) AS raw_invoices
FROM retail_raw_v2
WHERE customerid IS NOT NULL;
raw_invoices
-----
      22190
(1 row)
```

```
app_db=# -- Zeilen im Raw mit CustomerID, aber ohne passende Rechnung
SELECT COUNT(*) AS zeilen_ohne_invoice
FROM retail_raw_v2 r
LEFT JOIN invoices i USING (invoiceno)
WHERE r.customerid IS NOT NULL
    AND i.invoiceno IS NULL;
zeilen_ohne_invoice
-----
          0
(1 row)
```

```
app_db=# SELECT COUNT(*) AS problemrows
FROM retail_raw_v2 r
LEFT JOIN customers c USING (customerid)
LEFT JOIN products p USING (stockcode)
LEFT JOIN invoices i USING (invoiceno)
WHERE r.customerid IS NOT NULL -- nur Bestellungen mit Kunde
    AND (
        c.customerid IS NULL
        OR p.stockcode IS NULL
        OR i.invoiceno IS NULL
    );
problemrows
-----
          0
(1 row)
```