-- EDA Questions and Queries
-- ---------------------------

-- 1. How many total records are in the dataset?
SELECT COUNT(*) AS Total_Records FROM cohort_analysis.RETAIL;

TOTAL_RECORDS
541909

-- 2. How many unique customers are there?
SELECT COUNT(DISTINCT CustomerID) AS Unique_Customers FROM
cohort_analysis.RETAIL WHERE CustomerID IS NOT NULL;

UNIQUE_CUSTOMERS
4372

-- 3. How many distinct countries are in the data?
SELECT DISTINCT Country FROM cohort_analysis.RETAIL limit 12;

	A COUNTRY
1	United Kingdom
2	France
3	Australia
4	Netherlands
5	Germany
6	Norway
7	EIRE
8	Switzerland
9	Spain
10	Poland
11	Portugal
12	Belgium

-- 4. What is the range of dates in the data?
SELECT
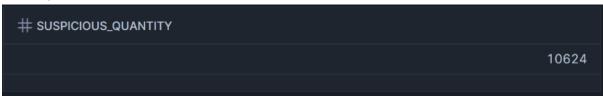   MIN(TO_TIMESTAMP(InvoiceDate, 'MM/DD/YYYY HH24:MI')) AS Start_Date,
    MAX(TO_TIMESTAMP(InvoiceDate, 'MM/DD/YYYY HH24:MI')) AS End_Date
FROM cohort_analysis.RETAIL;

⏲ START_DATE	⏲ END_DATE
2010-12-01 08:26:00.000	2011-12-09 12:50:00.000

-- 5. Are there any negative or zero quantity values?
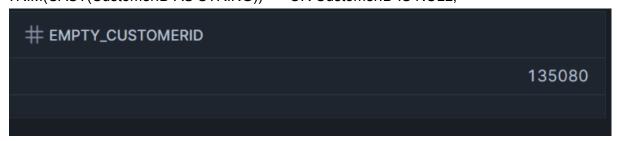SELECT COUNT(*) AS Suspicious_Quantity FROM cohort_analysis.RETAIL WHERE
Quantity <= 0;

```
# SUSPICIOUS_QUANTITY

                                                    10624
```

-- 6. Are there any negative or zero unit prices?
SELECT COUNT(*) AS Suspicious_UnitPrice FROM cohort_analysis.RETAIL WHERE
UnitPrice <= 0;

```
# SUSPICIOUS_UNITPRICE

                                                    2521
```

-- 7. How many canceled invoices (start with 'C')?
SELECT COUNT(*) AS Canceled_Orders FROM cohort_analysis.RETAIL WHERE
InvoiceNo LIKE 'C%';

```
# CANCELED_ORDERS

                                                    9288
```

-- 8. How many missing or empty Customer IDs?
SELECT COUNT(*) AS Empty_CustomerID FROM cohort_analysis.RETAIL WHERE
TRIM(CAST(CustomerID AS STRING)) = '' OR CustomerID IS NULL;

```
# EMPTY_CUSTOMERID

                                                    135080
```

-- 9. What are the top-selling products?
SELECT Description, SUM(Quantity) AS Total_Quantity

```
FROM cohort_analysis.RETAIL
WHERE Quantity > 0
GROUP BY Description
ORDER BY Total_Quantity DESC
LIMIT 10;
```

DESCRIPTION	TOTAL_QUANTITY
1 PAPER CRAFT , LITTLE BIRDIE	80995.00
2 MEDIUM CERAMIC TOP STORAGE JAR	78033.00
3 WORLD WAR 2 GLIDERS ASSTD DESIGNS	55047.00
4 JUMBO BAG RED RETROSPOT	48478.00
5 WHITE HANGING HEART T-LIGHT HOLDER	37895.00

```
-- 10. Which countries generate the most revenue?
SELECT Country, ROUND(SUM(Quantity * UnitPrice), 2) AS Revenue
FROM cohort_analysis.RETAIL
GROUP BY Country
ORDER BY Revenue DESC LIMIT  5;
```

COUNTRY	REVENUE
1 United Kingdom	8187806.36
2 Netherlands	284661.54
3 EIRE	263276.82
4 Germany	221698.21
5 France	197403.90

```
-- 11. Revenue trend over time
SELECT
    TO_DATE(TO_TIMESTAMP(InvoiceDate, 'MM/DD/YYYY HH24:MI')) AS Sale_Date,
    ROUND(SUM(Quantity * UnitPrice), 2) AS Daily_Revenue
FROM cohort_analysis.RETAIL
GROUP BY Sale_Date
ORDER BY Sale_Date;
```

2010-12-01	58635.56
2010-12-02	46207.28
2010-12-03	45620.46
2010-12-05	31383.95
2010-12-06	53860.18
2010-12-07	45059.05

-- 12. Average basket size per order
SELECT
    InvoiceNo,
    ROUND(SUM(Quantity * UnitPrice), 2) AS Order_Value,
    COUNT(DISTINCT StockCode) AS Product_Count
FROM cohort_analysis.RETAIL
GROUP BY InvoiceNo;

	A INVOICENO	# ORDER_VALUE	# PRODUCT_COUNT
1	536365	139.12	7
2	536366	22.20	2
3	536367	278.73	12
4	536368	70.05	4
5	536369	17.85	1
6	536370	855.86	20