What is the total number of records in dataset?

```
SELECT count(*) FROM sales;
```

	Records
▸	9694

What is the overall total sales, quantity sold, and profit?

```sql
SELECT
    ROUND(SUM(sales), 0) AS Total_Sales,
    ROUND(SUM(Quantity), 0) AS Quantity_Sold,
    ROUND(SUM(profit), 0) AS Profit
FROM
    sales;
```

	Total_Sales	Quantity_Sold	Profit
▸	2272450	36749	282858

How many orders were placed in each year?

```sql
SELECT
    YEAR(Order_Date) AS 'Year',
    COUNT(DISTINCT Order_ID) AS Orders
FROM
    sales
GROUP BY YEAR(Order_Date)
ORDER BY YEAR(Order_Date);
```

Year	Orders
2014	958
2015	1020
2016	1293
2017	1660

What is the average sales per month?

```sql
SELECT
    DATE_FORMAT(Order_Date, '%Y-%m') AS Month,
    ROUND(AVG(sales), 0) AS Average_Sales
FROM
    sales
GROUP BY Month
ORDER BY Month;
SET GLOBAL sql_mode = 'ANSI_QUOTES';
```

Month	Average_Sales
2014-01	191
2014-02	96
2014-03	361
2014-04	213
2014-05	195
2014-06	259
2014-07	245
2014-08	184

What is the earliest and latest order date in the dataset?

```sql
SELECT
    MIN(Order_Date) AS earliest_date,
    MAX(Order_Date) AS latest_date
FROM
    sales;
```

earliest_date	latest_date
2014-01-04	2017-12-30

How many orders were placed in each year?

```sql
SELECT
    DATE_FORMAT(Order_Date, '%Y') AS Year,
    COUNT(DISTINCT Order_ID) AS Orders
FROM
    sales
GROUP BY Year
ORDER BY Year;
```

Year	Orders
2014	958
2015	1020
2016	1293
2017	1660

What are the top 5 cities with highest sales?

```sql
SELECT
    City, ROUND(SUM(Sales), 0) AS Total_Sales
FROM
    sales
GROUP BY City
ORDER BY Total_Sales DESC
LIMIT 5;
```

City	Total_Sales
New York City	255249
Los Angeles	173169
Seattle	117773
San Francisco	110917
Philadelphia	107486

Which region has the highest avg profit margin?

```sql
SELECT
    Region, ROUND(AVG(Profit), 0) AS Average_Profit
FROM
    sales
GROUP BY Region
ORDER BY Average_Profit DESC
LIMIT 1;
```

Region	Average_Profit
West	34

Find the distribution of sales across different states

```sql
SELECT
    State, ROUND(SUM(sales), 0) AS Total_Sales
FROM
    sales
GROUP BY State;
```

State	Total_Sales
Kentucky	36592
California	450568
Florida	88877
North Carolina	55548
Washington	136590
Texas	169554
Wisconsin	32092

Identify the top 10 best-selling products by quantity and sales.

```sql
SELECT
    Product, SUM(Quantity) AS Quantity
FROM
    sales
GROUP BY Product
ORDER BY Quantity DESC
LIMIT 10;
```

Product	Quantity
Staples	215
Staple envelope	170
Easy-staple paper	150
Staples in misc. colors	86
KI Adjustable-Height Table	74
Storex Dura Pro Binders	71
Avery Non-Stick Binders	71

```sql
SELECT
    Product, ROUND(SUM(Sales), 0) AS Sales
FROM
    sales
GROUP BY Product
ORDER BY Sales DESC
LIMIT 10;
```

Product	Sales
Canon imageCLASS 2200 Advanced Copier	61600
Fellowes PB500 Electric Punch Plastic Comb Binding Machine ...	27453
Cisco TelePresence System EX90 Videoconferencing Unit	22638
HON 5400 Series Task Chairs for Big and Tall	21871
GBC DocuBind TL300 Electric Binding System	19823
GBC Ibimaster 500 Manual ProClick Binding System	19024
Hewlett Packard LaserJet 3310 Copier	18840

What is the total sales and quantity sold for each product category and sub-category?

```sql
SELECT
    Category,
    ROUND(SUM(Sales),0) AS Total_Sales,
    SUM(Quantity) AS Total_Quantity
FROM
    sales
GROUP BY Category;
```

```sql
SELECT
    Sub_Category,
    ROUND(SUM(Sales),0) AS Total_Sales,
    SUM(Quantity) AS Total_Quantity
FROM
    sales
GROUP BY Sub_Category;
```

Category	Total_Sales	Total_Quantity
Furniture	733047	7855
Office Supplies	703503	21990
Technology	835900	6904

Sub_Category	Total_Sales	Total_Quantity
Bookcases	114880	868
Chairs	328449	2356
Labels	12486	1400
Tables	206966	1241
Storage	216803	3065
Furnishings	82752	3390
Art	27119	3000

Determine the products with the highest profit margins.

```sql
SELECT
    Product, ROUND(SUM(Profit), 0) AS Profit
FROM
    sales
GROUP BY Product
ORDER BY Profit DESC
LIMIT 3;
```

Product	Profit
Canon imageCLASS 2200 Advanced Copier	25200
Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind	7753
Hewlett Packard LaserJet 3310 Copier	6984

Determine the products with the lowest profit margins.

```sql
SELECT
    Product, ROUND(SUM(Profit), 0) AS Profit
FROM
    sales
GROUP BY Product
ORDER BY Profit ASC
LIMIT 3;
```

Product	Profit
Cubify CubeX 3D Printer Double Head Print	-8880
Lexmark MX611dhe Monochrome Laser Printer	-4590
Cubify CubeX 3D Printer Triple Head Print	-3840

Who are the top 10 customers based on total sales?

```sql
SELECT
    Customer, ROUND(SUM(Sales), 0) AS Sales
FROM
    sales
GROUP BY Customer
ORDER BY Sales DESC
LIMIT 10;
```

Customer	Sales
Sean Miller	25043
Tamara Chand	19018
Raymond Buch	15117
Tom Ashbrook	14596
Adrian Barton	14356
Sanjit Chand	14142
Ken Lonsdale	14072
Hunter Lopez	12873
Sanjit Engle	12209
Christopher Conant	12129

Identify the customers who made the most and least purchases.

```sql
SELECT
    Customer, SUM(quantity) AS Most_Purchases
FROM
    sales
GROUP BY Customer
ORDER BY Most_Purchases DESC
LIMIT 1;
```

```sql
SELECT
    Customer, SUM(quantity) AS Least_Purchases
FROM
    sales
GROUP BY Customer
ORDER BY Least_Purchases ASC
LIMIT 1;
```

Customer	Most_Purchases
John Lee	138

Customer	Least_Purchases
Anthony O'Donnell	2

What is the average discount given to customers in each segment?

```sql
SELECT
    Segment, CONCAT(ROUND(AVG(Discount), 3)*100,'%') AS Discount
FROM
    sales
GROUP BY Segment;
```

Segment	Discount
Consumer	15.7%
Corporate	15.7%
Home Office	14.7%

Find the most common shipping mode.

```sql
SELECT
    Ship_Mode, COUNT(*) AS count
FROM
    sales
GROUP BY Ship_Mode;
```

Ship_Mode	count
Second Class	1886
Standard Class	5780
First Class	1501
Same Day	527

Calculate the average shipping time (difference between order and ship dates).

```sql
SELECT
    ROUND(AVG(DATEDIFF(Ship_Date, Order_Date)),0) AS Avg_Shipping_Days
FROM
    sales
WHERE
    Order_Date IS NOT NULL
        AND Ship_Date IS NOT NULL;
```

Avg_Shipping_Days
4

Identify any trends or patterns in delayed shipments.

```sql
ALTER TABLE sales
ADD COLUMN Duration INT;
UPDATE sales
SET
    Duration = DATEDIFF(Ship_Date, Order_Date);

SELECT
    Region, COUNT(DISTINCT Order_ID) AS Number_of_Late_Orders
FROM
    sales
WHERE
    Duration > (SELECT
            AVG(DATEDIFF(Ship_Date, Order_Date))
        FROM
            sales)
GROUP BY Region
ORDER BY Number_of_Late_Orders DESC;
```

Region	Number_of_Late_Orders
West	1052
East	920
Central	812
South	552

Investigate the relationship between discounts and profit.

```sql
SELECT
    Discount, ROUND(AVG(Profit), 0) AS Profit
FROM
    sales
GROUP BY Discount
ORDER BY Discount;
```

Discount	Profit
0	68
0.1	96
0.15	27
0.2	25
0.3	-46
0.32	-89
0.4	-113
0.45	-227
0.5	-311
0.6	-42

Determine the impact of discounts on sales and quantity.

```sql
SELECT
    Discount,
    ROUND(AVG(Sales), 2) AS Avg_Sales,
    ROUND(AVG(Quantity), 2) AS Avg_Quantity
FROM
    sales
GROUP BY Discount;
```

Discount	Avg_Sales	Avg_Quantity
0	230.36	3.81
0.1	578.4	3.97
0.15	529.97	3.81
0.2	213.22	3.74
0.3	454.74	3.74
0.32	536.79	3.89
0.4	567.54	3.78
0.45	498.63	4.09
0.5	892.71	3.65

What is the distribution of sales and profit across different segments?

```sql
SELECT
    Segment,
    ROUND(SUM(Sales), 0) AS Avg_sales,
    ROUND(SUM(Profit), 0) AS Avg_Profit
FROM
    sales
GROUP BY Segment;
```

Segment	Avg_sales	Avg_Profit
Consumer	1150166	132670
Corporate	696605	90366
Home Office	425679	59822

What is the distribution of sales and profit across different segments?

```sql
SELECT
    Segment,
    ROUND(AVG(Sales), 0) AS Avg_sales,
    ROUND(AVG(Profit), 0) AS Avg_Profit
FROM
    sales
GROUP BY Segment;
```

Segment	Avg_sales	Avg_Profit
Consumer	228	26
Corporate	239	31
Home Office	246	35

Calculate the correlation between sales, quantity, discount, and profit.

```
SELECT
    ROUND((COUNT(*) * SUM(Sales * Profit) - SUM(Sales) * SUM(Profit)) / (SQRT((COUNT(*) * SUM(Sales * Sales) - POW(SUM(Sales), 2)) * (COUNT(*) * SUM(Profit * Profit) - POW(SUM(Profit), 2)))),
            2) AS Sales_Profit_Correlation FROM sales;
```

Sales_Profit_Correlation
0.48

```
   SELECT
    ROUND((COUNT(*) * SUM(Quantity * Profit) - SUM(Quantity) * SUM(Profit)) / (SQRT((COUNT(*) * SUM(Quantity * Quantity) - POW(SUM(Quantity), 2)) * (COUNT(*) * SUM(Profit * Profit) - POW(SUM(Profit), 2)))),
            2) AS Quantity_Profit_Correlation FROM sales;
```

Quantity_Profit_Correlati
0.07

```
 SELECT
    ROUND((COUNT(*) * SUM(Discount * Profit) - SUM(Discount) * SUM(Profit)) / (SQRT((COUNT(*) * SUM(Discount * Discount) - POW(SUM(Discount), 2)) * (COUNT(*) * SUM(Profit * Profit) - POW(SUM(Profit), 2)))),
            2) AS Discount_Profit_Correlation FROM sales;
```

Quantity_Profit_Correlation
0.07