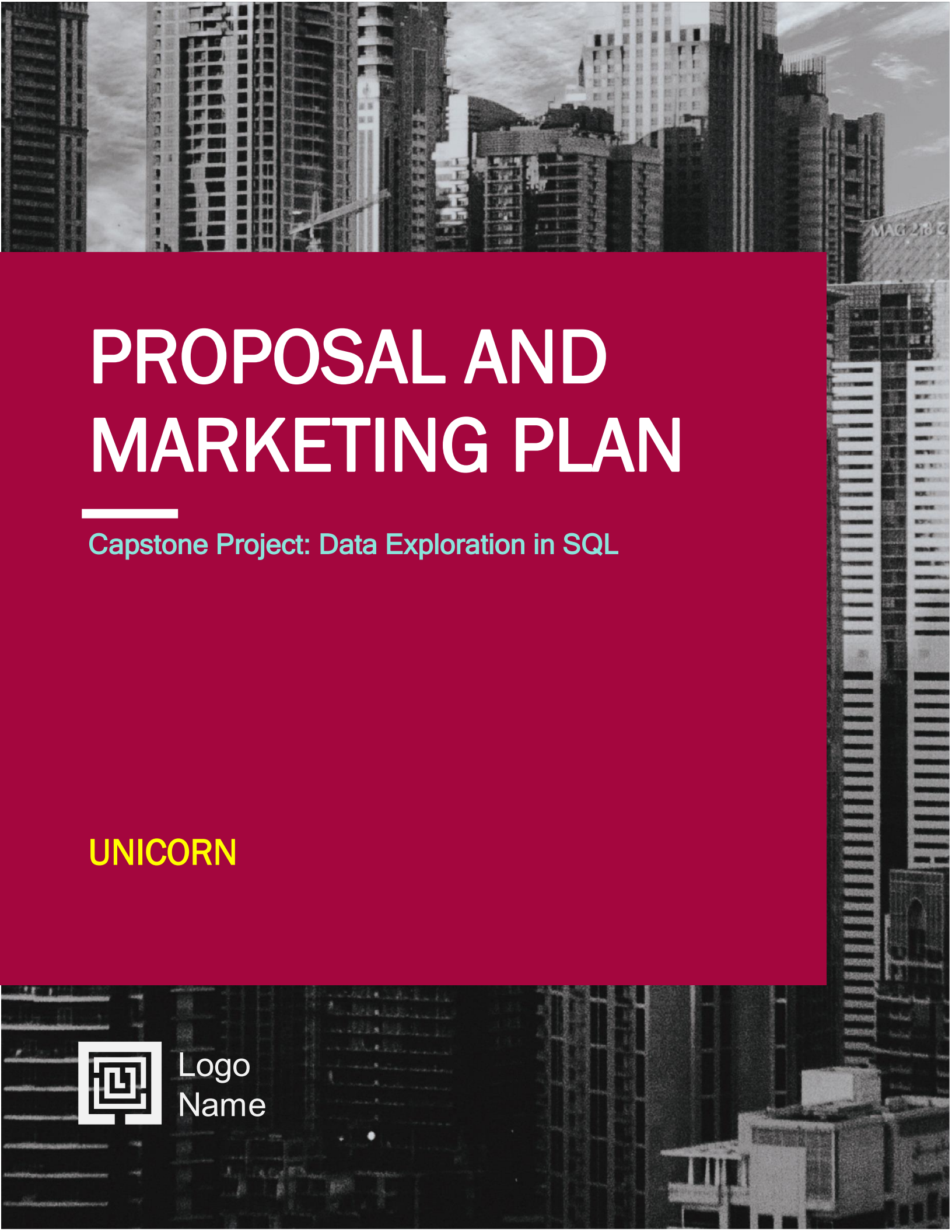


PROPOSAL AND MARKETING PLAN

Capstone Project: Data Exploration in SQL

UNICORN

Logo Name

List of columns per table

customers ◦ customer_id ◦ customer_name ◦ customer_segment	**product** ◦ product_id ◦ product_name ◦ product_category ◦ product_subcategory ◦ product_manufacturer
order_details ◦ order_details_id ◦ order_id ◦ product_id ◦ quantity ◦ order_discount ◦ order_profits ◦ order_profit_ratio ◦ order_sales	**orders** ◦ order_id ◦ customer_id ◦ order_date ◦ shipping_city ◦ shipping_state ◦ shipping_region ◦ shipping_country ◦ shipping_postal_code ◦ shipping_date shipping_mode

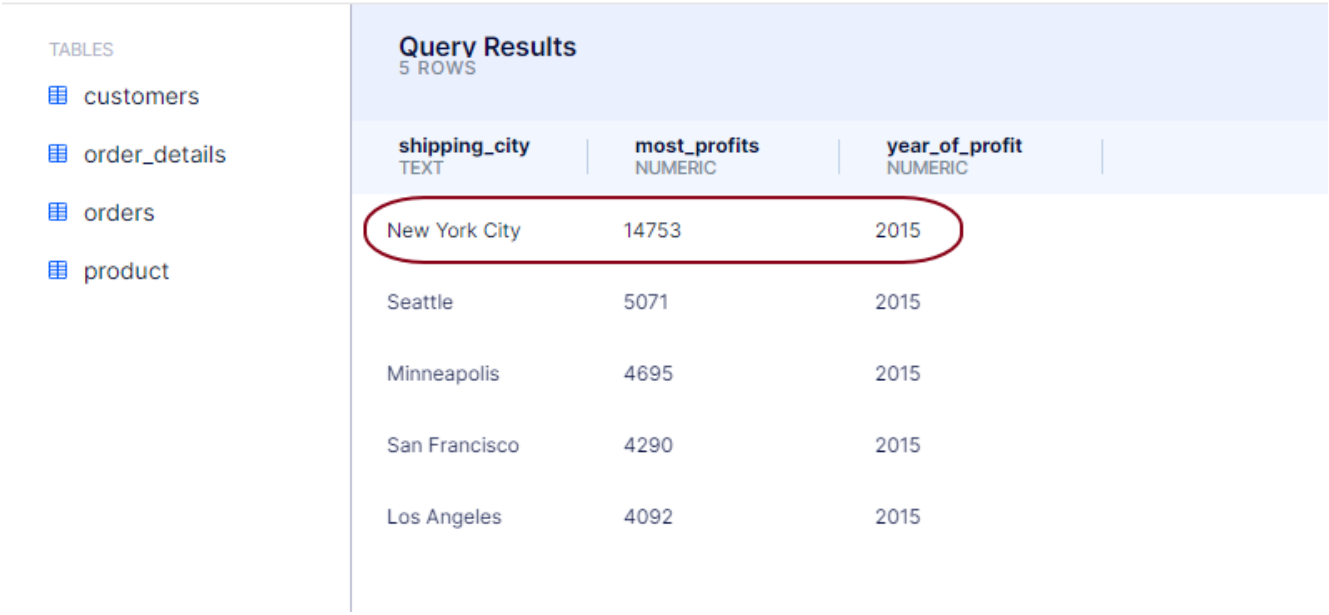
Questions:

1. **How many customers do we have in the data?**

```
SELECT Count (*)
FROM customers
```

2. **What was the city with the most profit for the company in 2015?**

```
SELECT shipping_city,SUM(order_profits),EXTRACT(YEAR FROM order_date) AS Year_of_profit
FROM Orders AS o
JOIN order_details AS od
ON o.order_id = od.order_id
WHERE EXTRACT(YEAR FROM order_date) = '2015'
GROUP BY 1,3
ORDER BY 2 DESC
LIMIT 5;
```



3. In 2015, what was the most profitable city's profit?

```
SELECT shipping_city,SUM(order_profits),EXTRACT(YEAR FROM order_date) AS Year_of_profit
FROM Orders AS o
JOIN order_details AS od
ON o.order_id = od.order_id
WHERE EXTRACT(YEAR FROM order_date) = '2015'
GROUP BY 1,3
ORDER BY 2 DESC
LIMIT 5;
```

TABLES
- customers
- order_details
- orders
- product

Query Results
5 ROWS

shipping_city (TEXT)	most_profits (NUMERIC)	year_of_profit (NUMERIC)
New York City	14753	2015
Seattle	5071	2015
Minneapolis	4695	2015
San Francisco	4290	2015
Los Angeles	4092	2015

4. How many different cities do we have in the data?

```
SELECT  COUNT (DISTINCT shipping_city) AS different_cities_nos
FROM orders
```

TABLES
- customers
- order_details
- orders
- product

Query Results
1 ROWS

different_cities_nos (INT8)
531

5. Show the total spent by customers from low to high.

```sql
SELECT c.customer_id,c.customer_name,SUM(od.order_sales) AS total_spent
FROM order_details AS od
JOIN orders AS o
ON o.order_id = od.order_id
JOIN customers AS c
ON c.customer_id = o.customer_id
GROUP BY 1,2
ORDER BY 3
LIMIT 10;
```

TABLES

- customers
- order_details
- orders
- product

Query Results
10 ROWS

customer_id INT8	customer_name TEXT	total_spent NUMERIC
456	Lela Donovan	5
738	Thais Sissman	5
546	Mitch Gastineau	16
124	Carl Jackson	17
657	Roy Skaria	22
626	Ricardo Emerson	48
725	Susan Gilcrest	49
448	Larry Blacks	50
9	Adrian Shami	58
355	Jasper Cacioppo	72

6. What is the most profitable city in the State of Tennessee?

```sql
SELECT o.shipping_city, SUM(od.order_profits) AS most_profit
FROM order_details AS od
JOIN orders AS o
ON o.order_id = od.order_id
WHERE o.shipping_state = 'Tennessee'
GROUP BY 1
ORDER BY SUM(od.order_profits) DESC
LIMIT 5;
```

TABLES

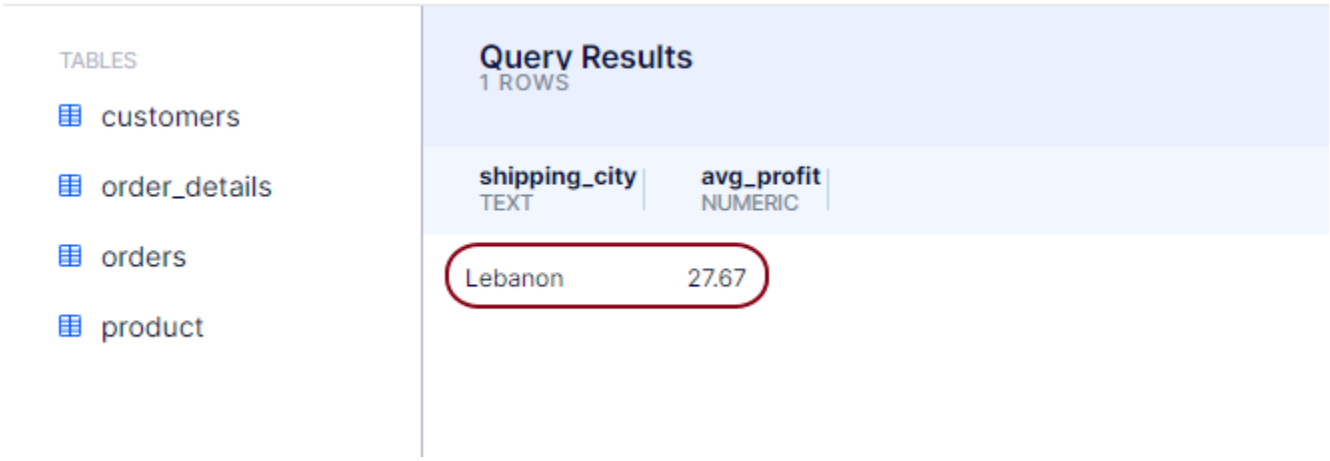
- customers
- order_details
- orders
- product

Query Results
5 ROWS

shipping_city TEXT	most_profit NUMERIC
Lebanon	83
Chattanooga	28
Smyrna	20
Murfreesboro	15
Johnson City	12

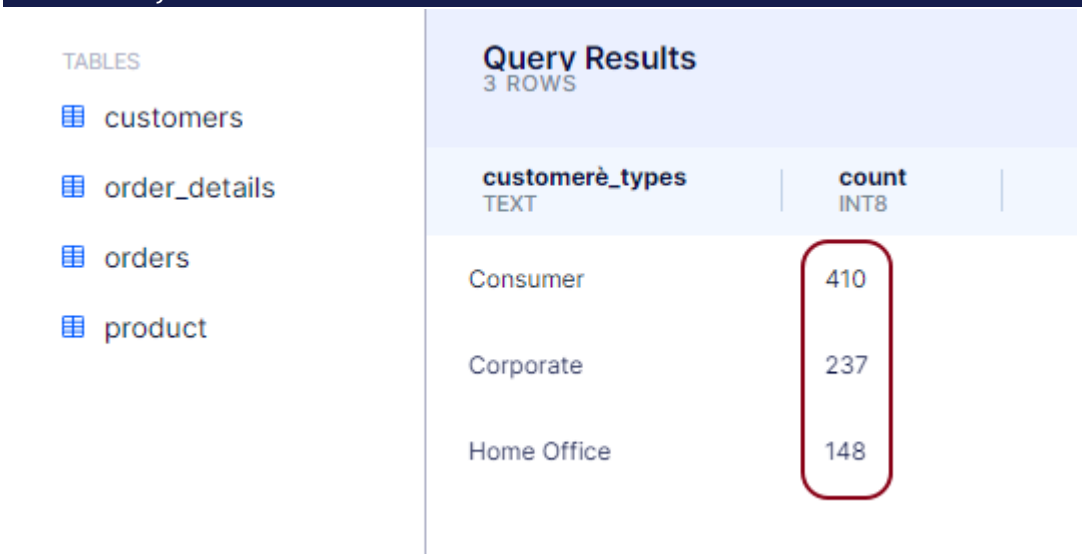
7. What's the average annual profit for that city across all years?

```sql
SELECT o.shipping_city, ROUND(AVG(od.order_profits),2) AS avg_profit
FROM order_details AS od
JOIN orders AS o
ON o.order_id = od.order_id
WHERE o.shipping_city = 'Lebanon'
GROUP BY 1
ORDER BY AVG(od.order_profits) DESC;
```



8. **What is the distribution of customer types in the data?**

```
SELECT DISTINCT customer_segment AS customerè_types, count(*)
FROM customers
GROUP BY 1;
```

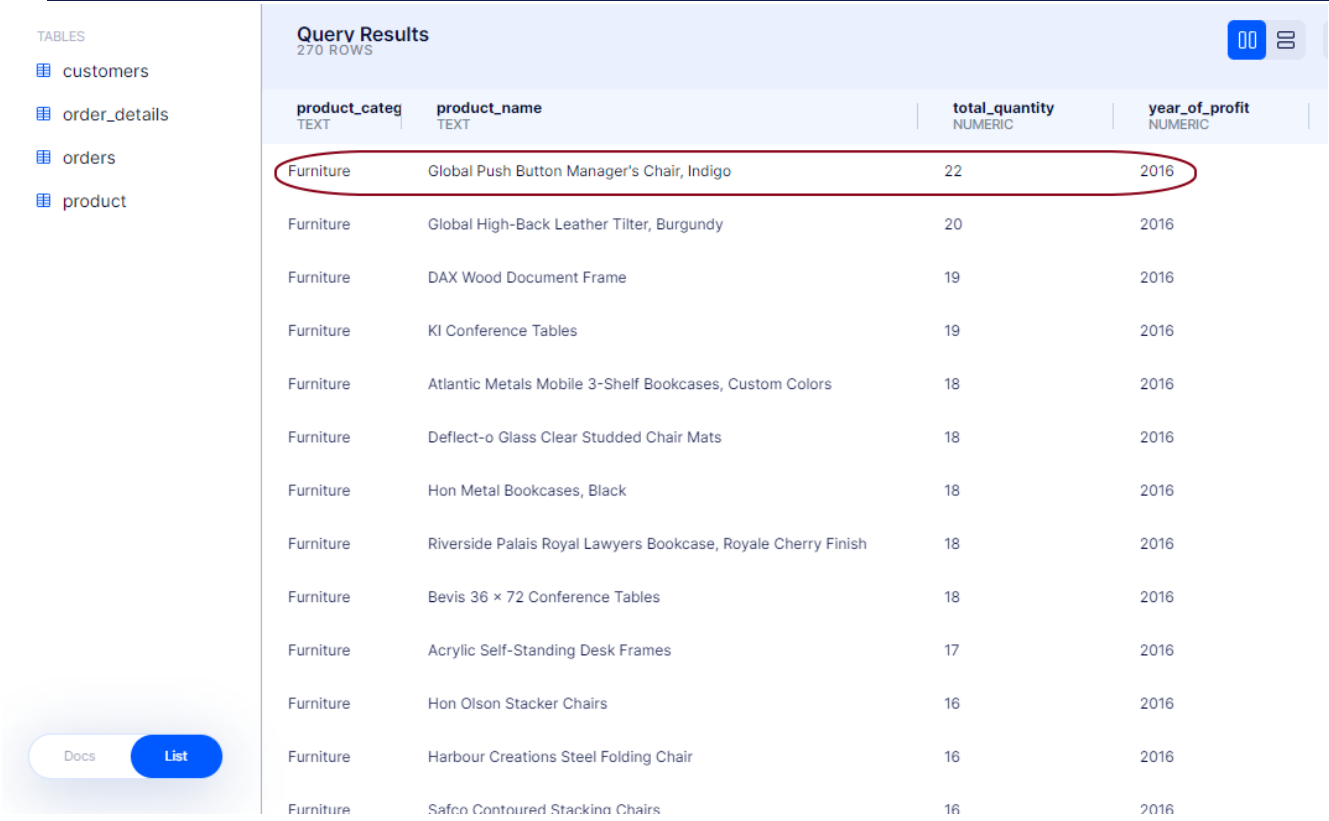


9. **What's the most profitable product category on average in Iowa across all years?**

```
SELECT p.product_category,o.shipping_state,AVG(od.order_profits) AS
avg_profitable_category
FROM product AS p
JOIN order_details AS od
ON p.product_id = od.product_id
JOIN orders AS o
ON o.order_id = od.order_id
WHERE o.shipping_state = 'Iowa'
GROUP BY 1,2
ORDER BY AVG(od.order_profits) DESC;
```

10. What is the most popular product in that category across all states in 2016?

```sql
SELECT p.product_category,p.product_name,SUM(od.quantity) AS total_quantity,
EXTRACT(YEAR FROM order_date) AS Year_of_profit
FROM product AS p
JOIN order_details AS od
ON p.product_id = od.product_id
JOIN orders AS o
ON o.order_id = od.order_id
WHERE EXTRACT (YEAR FROM o.order_date) = '2016' AND p.product_category = 'Furniture'
GROUP BY 1,2,4
ORDER BY SUM(od.quantity) DESC;
```



11. Which customer got the most discount in the data? (in total amount)

```
SELECT c.customer_id,c.customer_name, SUM((order_sales*order_discount)/(1-
order_discount)) AS total_discount
FROM customers AS c
JOIN orders AS o
ON c.customer_id = o.customer_id
JOIN order_details AS od
ON o.order_id = od.order_id
GROUP BY 1,2
ORDER BY 3 DESC
LIMIT 5;
```

TABLES

- customers
- order_details
- orders
- product

Query Results
5 ROWS

customer_id INT8	customer_name TEXT	total_discount FLOAT8
687	Sean Miller	23929.083333333332
166	Cindy Stewart	11594.333333333332
478	Luke Foster	9052.166666666668
308	Grant Thornton	8746.833333333334
330	Henry Goldwyn	7692.583333333335

12. How widely did monthly profits vary in 2018?

```
SELECT EXTRACT(MONTH FROM o.order_date) AS MONTH, SUM(od.order_profits) AS
monthly_profits,LAG (SUM(od.order_profits),1,0) OVER(ORDER BY EXTRACT(MONTH FROM
o.order_date))
AS lag, (SUM(od.order_profits)-LAG (SUM(od.order_profits),1,0) OVER(ORDER BY
EXTRACT(MONTH FROM o.order_date))) AS variance_monthly

FROM orders AS o
JOIN order_details AS od
ON o.order_id = od.order_id
WHERE EXTRACT(YEAR FROM order_date) = '2018'
GROUP BY 1
ORDER BY 1
```

TABLES
- customers
- order_details
- orders
- product

Docs | List

Query Results
12 ROWS

month NUMERIC	monthly_profits NUMERIC	lag NUMERIC	variance_monthly NUMERIC
1	7137	0	7137
2	1612	7137	-5525
3	14758	1612	13146
4	934	14758	-13824
5	6342	934	5408
6	8226	6342	1884
7	6951	8226	-1275
8	9034	6951	2083
9	10987	9034	1953
10	9272	10987	-1715
11	9217	9272	-55
12	8473	9217	-744

13. Which order was the highest in 2015?

```sql
SELECT p.product_subcategory,od.order_id,SUM(od.order_sales) AS total_sales
FROM orders AS o
JOIN order_details AS od
ON o.order_id = od.order_id
JOIN product AS p
ON p.product_id = od.product_id
WHERE EXTRACT(YEAR FROM order_date) = '2015'
GROUP BY 1,2
ORDER BY 3 DESC
LIMIT 1;
```

TABLES
- customers
- order_details
- orders
- product

Query Results
1 ROWS

product_subca TEXT	order_id TEXT	sum NUMERIC
Machines	CA-2015-145317	23459

14. What was the rank of each city in the East region in 2015?

```
SELECT o.shipping_city,
SUM(od.quantity) AS total_quantity, rank() OVER (Order by SUM(od.quantity)DESC) AS
city_rank
FROM orders AS o
JOIN order_details AS od
ON o.order_id = od.order_id
WHERE EXTRACT(YEAR FROM order_date) = '2015' AND o.shipping_region = 'East'
GROUP BY 1
ORDER BY 3;
```

TABLES

- customers
- order_details
- orders
- product

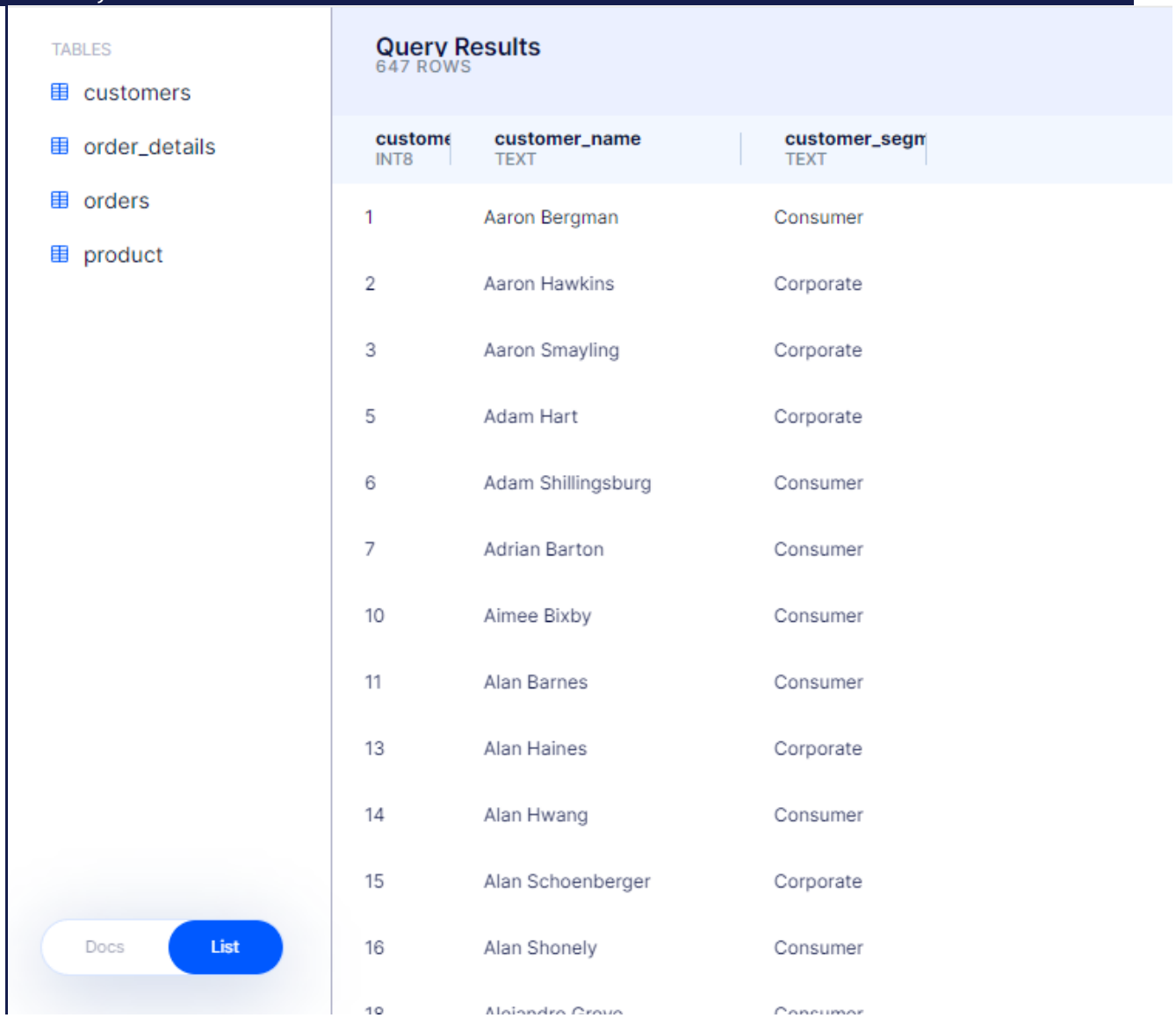
Docs | List

Query Results
60 ROWS

shipping_city TEXT	total_quantity NUMERIC	city_rank INT8
New York City	1708	1
Philadelphia	403	2
Columbus	167	3
Newark	64	4
Fairfield	53	5
Long Beach	44	6
Lakewood	38	7
Lancaster	36	8
Lawrence	31	9
Dover	30	10
Hackensack	28	11
Bangor	24	12
Utica	24	12

15. **Display customer names for customers who are in the segment 'Consumer' or 'Corporate.' How many customers are there in total?**

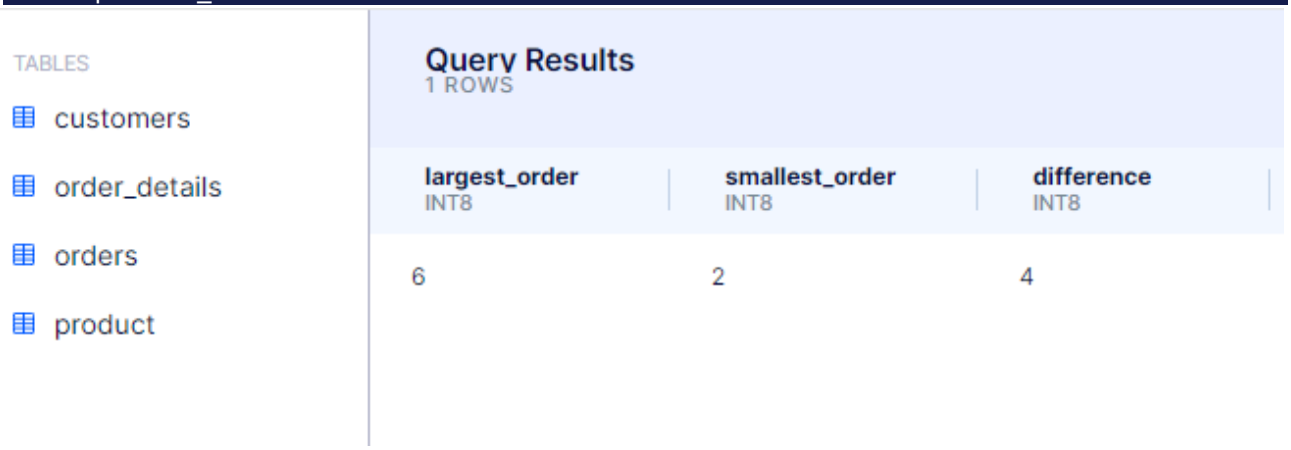
```
SELECT customer_id,customer_name,COUNT(*)
FROM customers
WHERE customer_segment IN ('Consumer', 'Corporate')
GROUP BY 1,2
```




```
SELECT COUNT(*)
FROM customers
WHERE
customer_segment IN ('Consumer', 'Corporate');
```

16. Calculate the difference between the largest and smallest order quantities for product id '100.'

```
select MAX(quantity) AS largest_order,MIN(quantity) AS smallest_order,
(MAX(quantity)-MIN(quantity)) AS difference
FROM order_details
WHERE product_id = '100'
```



17. Calculate the percent of products that are within the category 'Furniture.'

```
WITH t1 AS
(SELECT product_category,COUNT(*) AS count_category
FROM product
GROUP BY 1),
t2 AS
(SELECT COUNT(*) AS total_category
FROM product)

SELECT product_category, ROUND(t1.count_category*1.0/t2.total_category*100,2) AS
percentage_category
FROM t1,t2
```

```
OR
select
(SELECT COUNT(*) AS count_category FROM product where product_category = 'Furniture')*1.0/
(SELECT COUNT(*) AS count_category FROM product) as Furniture_percentage
```

18. **Display the number of duplicate products based on their product manufacturer. Example: A product with an identical product manufacturer can be considered a duplicate.**

```
select product_manufacturer,COUNT(*) AS num_of_duplicates
FROM product
GROUP BY 1
HAVING count(*)>1
```

TABLES

- customers
- order_details
- orders
- product

Docs | List

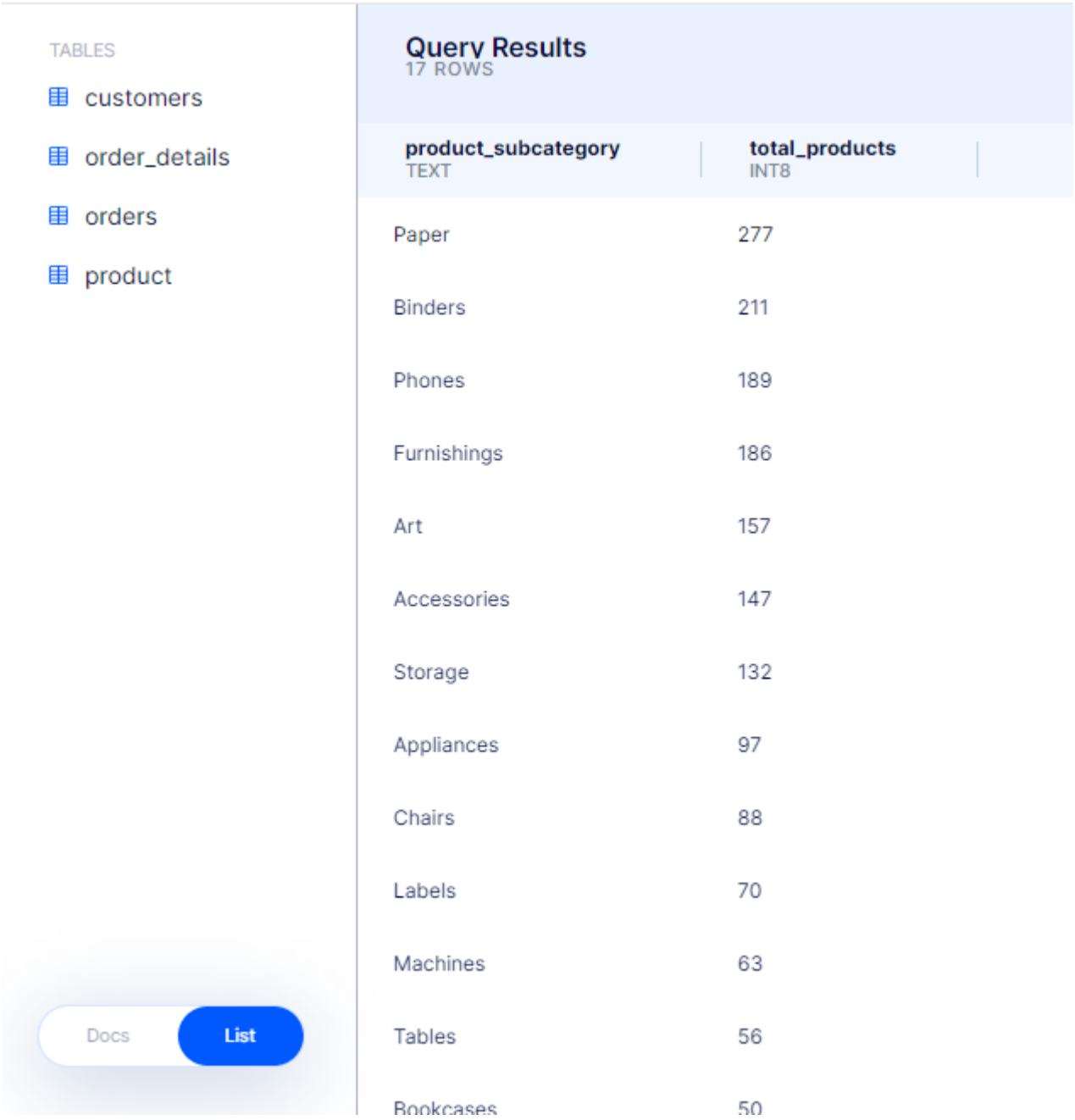
Query Results

169 ROWS

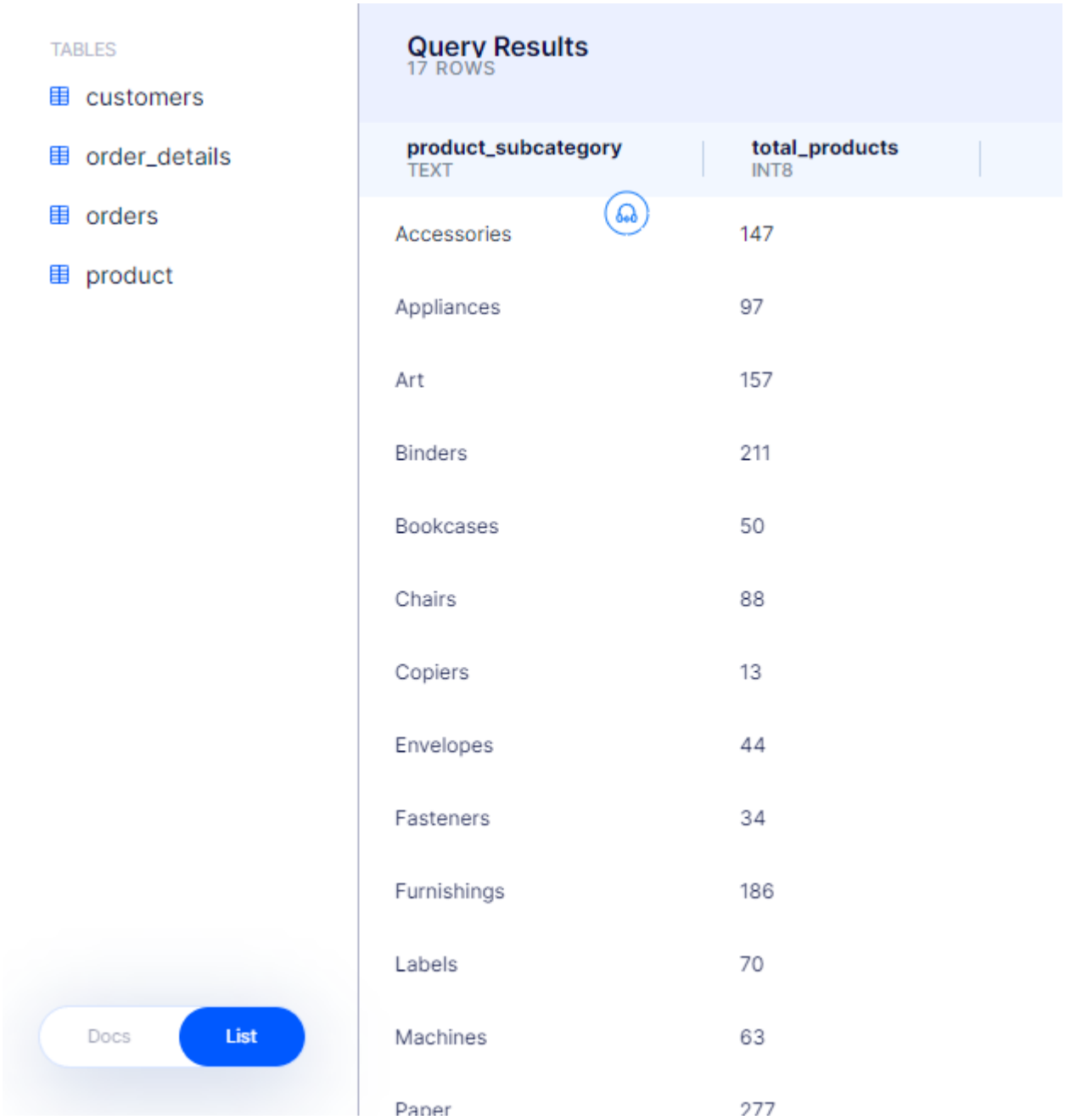
product_manufacturer TEXT	num_of_duplicates INT8
Linden	2
Iceberg	3
SanDisk	8
Memorex	13
Bulldog	2
Nortel	4
Quartet	2
Tyvek	4
Epson	3
XtraLife	2
Mitel	2
Grandstream	2
Panasonic	18

19. Show the product_subcategory and the total number of products in the subcategory. Show the order from most to least products and then by product_subcategory name ascending.

```
SELECT product_subcategory,COUNT(product_subcategory) AS total_products
FROM product
GROUP BY 1
ORDER BY 2 DESC;
```




```
SELECT product_subcategory,COUNT(product_subcategory) AS total_products
FROM product
GROUP BY 1
ORDER BY 1;
```



20. **Show the product_id(s), the sum of quantities, where the total sum of its product quantities is greater than or equal to 100.**

```
SELECT product_id,SUM(quantity) AS total_quantity
FROM order_details
GROUP BY 1
HAVING SUM(quantity) >= '100'
```

TABLES

- customers
- order_details
- orders
- product

Query Results
8 ROWS

product_id INT8	total_quantity NUMERIC
538	150
812	109
1216	132
1507	539
1501	170
920	155
1600	221
122	295

21 Join all database tables into one dataset that includes all unique columns and download it as a .csv file

```
SELECT *
FROM order_details
NATURAL JOIN orders
NATURAL JOIN customers
NATURAL JOIN product
```