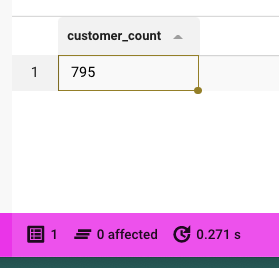
Unicorn Project - SQL Questions & Answers

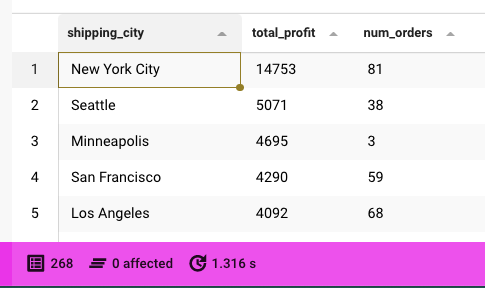
1. How many customers do we have in the data?

-- There are 795 customers in the database

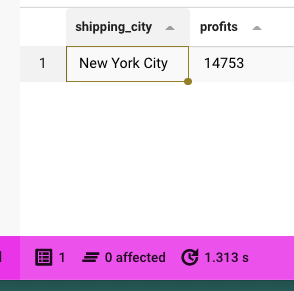


1. What was the city with the most profit for the company in 2015?

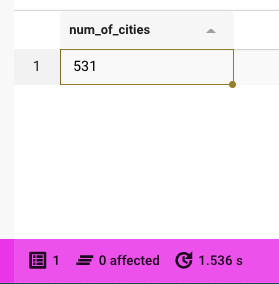
-- **ANS -- New York City with 14753**



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

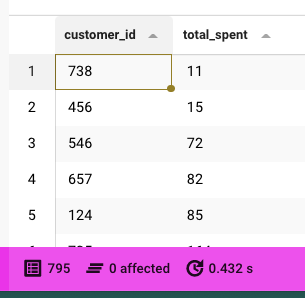


1. How many different cities do we have in the data?



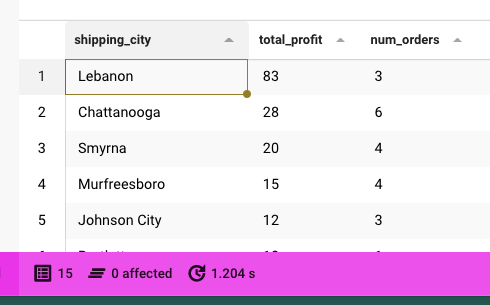
1. Show the total spent by customers from low to high.

-- The lowest amount spent is 11 and the highest amount spent is 1,1320,067

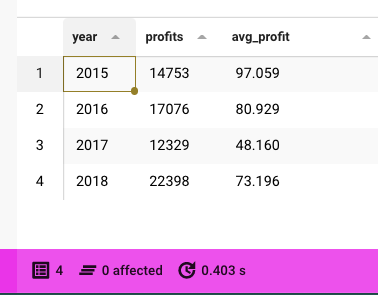


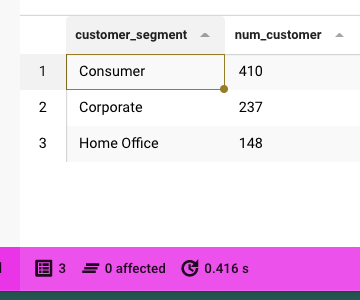
1. What is the most profitable city in the State of Tennessee?

**--ANS-- Lebanon is the most profitable city with $83 profit and 3 orders**

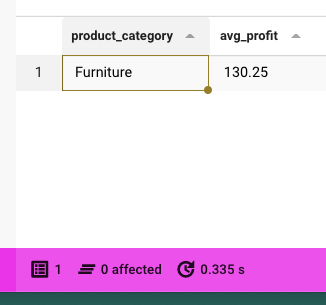


1. What’s the average annual profit for that city across all years?

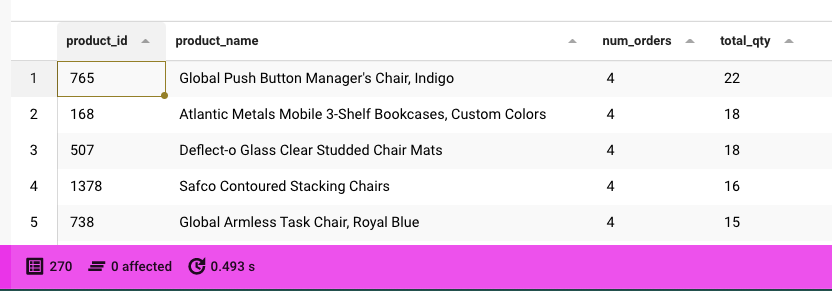


1. What is the distribution of customer types in the data?
2. What’s the most profitable product category on average in Iowa across all years?

-- The most profitable product category in Iowa is Furniture with an average profit of 130.25 and the least profitable was office supply with 15.73



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

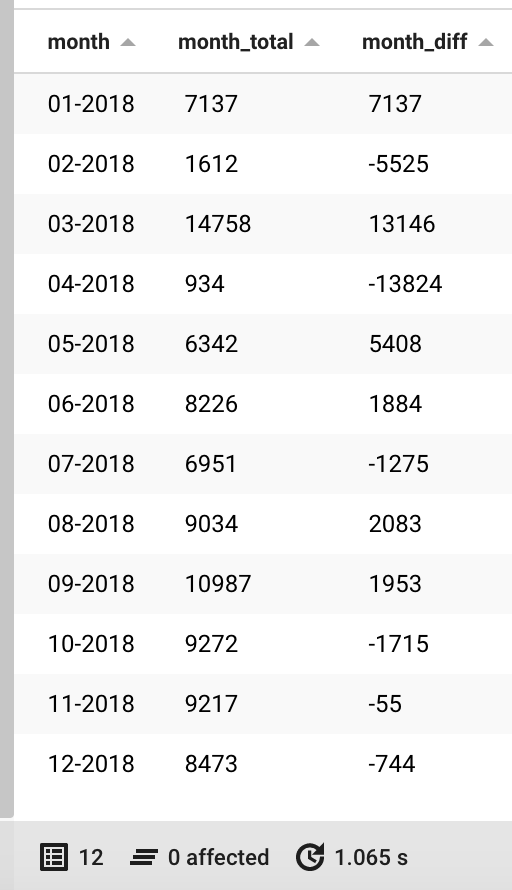
--The most popular product in 'Furniture' Category is Global Push Button Manager's Chair', indigo with 22 qty in 2016

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

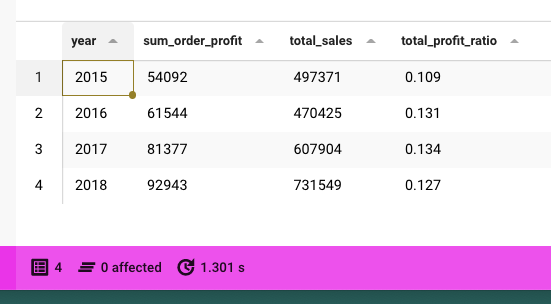
Ordered by: Total discount = order\_discount \* order\_sales

Ordered by: SUM(Total\_order\_discount)



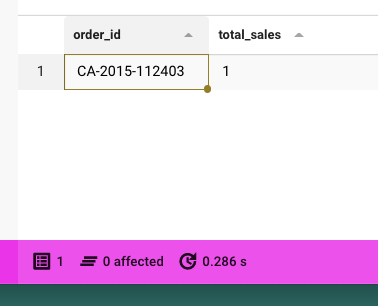
1. How widely did monthly profits vary in 2018? 

# This one is another code. I also added this yearly change to see how the profit ratio changed.

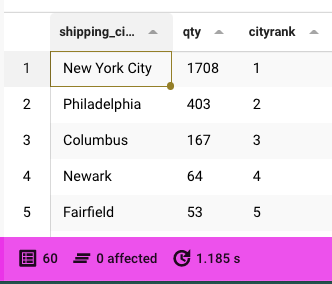


1. Which was the biggest order regarding sales in 2015?

**-- The biggest order sale was from order\_id CA-2015-145317 and was 23660. Least was CA-2015-112403 with 1 as sales price**

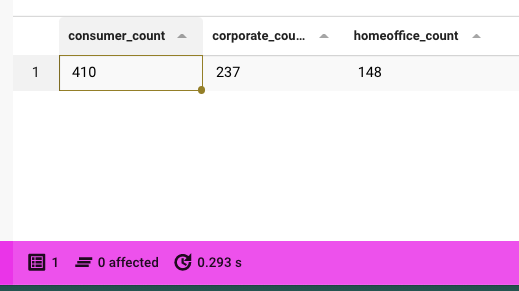


1. What was the rank of each city in the East region in 2015 in quantity?

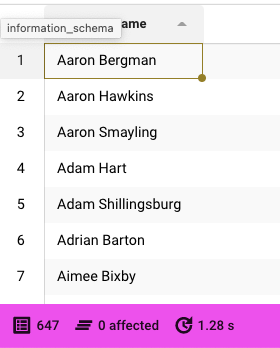


1. Display customer names for customers who are in the segment ‘Consumer’ or ‘Corporate.’ How many customers are there in total?

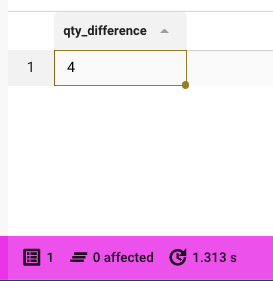
-- Unique customers by segment



-- List of Consumers or Corporate (alphabetic order first name)



1. Calculate the difference between the largest and smallest order quantities for product id ‘100.’



1. Calculate the percent of products that are within the category ‘Furniture.’

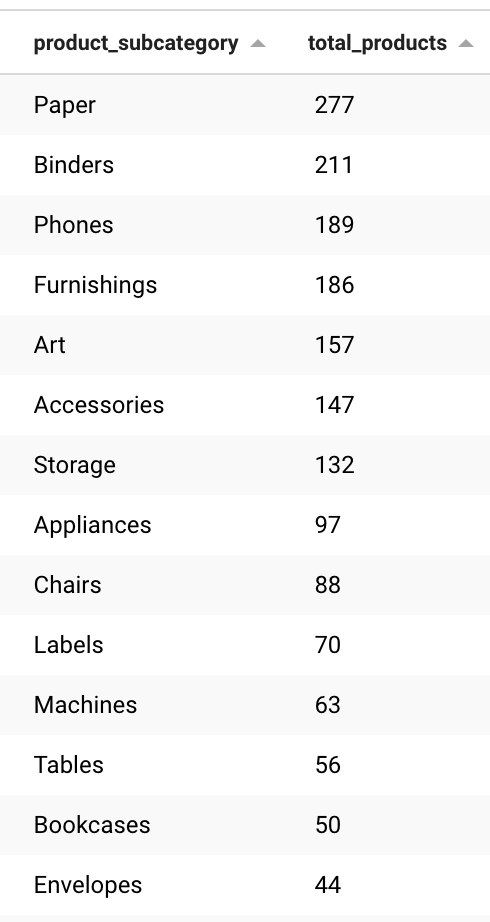
-- The percentage of Furniture products is 20.54%.



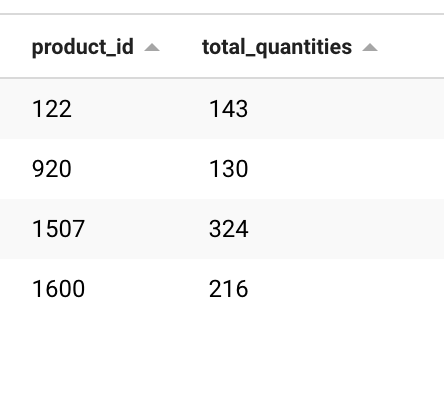
1. Display the number of product manufacturers with more than 1 product in the product table.



1. 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.



1. Show the product\_id(s), the sum of quantities, where the total sum of its product quantities is greater than or equal to 100.



**Bonus question:** Join all database tables into one dataset that includes all unique columns and download it as a .csv file.

WITH CTE\_ALL AS (

SELECT

o.\*, EXTRACT(YEAR FROM o.order\_date) AS year,

od.order\_details\_id, od.quantity, od.order\_discount,

od.order\_profits, od.order\_profit\_ratio, od.order\_sales,

c.customer\_name, c.customer\_segment,

p.\*

-- p.product\_id, p.product\_name, p.product\_category, p.product\_subcategory, p.product\_manufacturer

FROM orders o

JOIN order\_details od ON o.order\_id = od.order\_id

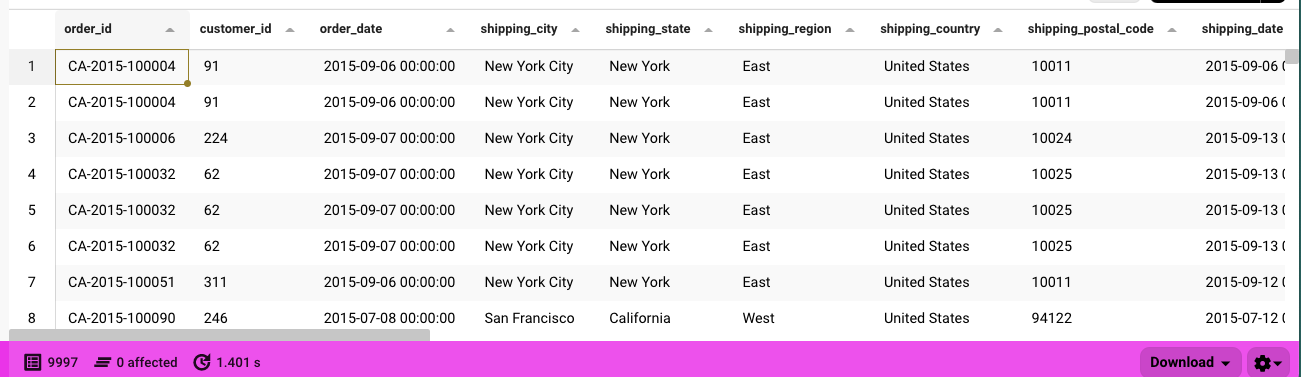
JOIN customers c ON o.customer\_id = c.customer\_id

JOIN product p ON p.product\_id = od.product\_id

)

SELECT \*

FROM CTE\_ALL

;

Supplementary Material

**Appendix 1: A list with the questions and the SQL code as used in Beekeeper.**

| **#** | **Question** | **SQL Code** |
| --- | --- | --- |
| 1 | How many customers do we have in the data? | SELECT  COUNT(customer\_id) AS customer\_count  FROM customers  ; |
| 2 | What was the city with the most profit for the company in 2015? | SELECT o.shipping\_city, SUM(od.order\_profits) AS Total\_profit,  COUNT(DISTINCT o.order\_id) AS Num\_orders  FROM orders o  LEFT JOIN order\_details od ON o.order\_id = od.order\_id  WHERE EXTRACT (YEAR FROM o.order\_date)=2015  GROUP BY 1  ORDER BY 2 DESC  Limit 3  ; |
| 3 | In 2015, what was the most profitable city's profit? | WITH CTE AS (  SELECT \*,  EXTRACT(YEAR FROM order\_date) AS year  FROM orders  JOIN order\_details ON orders.order\_id = order\_details.order\_id  )  -- SELECT \* FROM CTE  SELECT shipping\_city,  SUM(order\_profits) AS profits  FROM CTE  WHERE YEAR = 2015  GROUP BY shipping\_city  ORDER BY profits DESC  LIMIT 1; |
| 4 | How many different cities do we have in the data? | SELECT  COUNT(DISTINCT shipping\_city) AS Num\_of\_Cities  FROM orders  ; |
| 5 | Show the total spent by customers from low to high | SELECT cs.customer\_id,SUM(od.order\_sales \* od.quantity) total\_spent  FROM customers cs  LEFT JOIN orders o ON cs.customer\_id = o.customer\_id  LEFT JOIN order\_details od ON o.order\_id = od.order\_id  GROUP BY cs.customer\_id  ORDER BY total\_spent ; |
| 6 | What is the most profitable city in the State of Tennessee? | SELECT o.shipping\_city,  SUM(od.order\_profits) AS Total\_profit,  COUNT(DISTINCT o.order\_id) AS Num\_orders  FROM orders o  JOIN order\_details od ON o.order\_id = od.order\_id  WHERE shipping\_state LIKE '%Tenne%'  GROUP BY 1  ORDER BY 2 DESC  ; |
| 7 | What’s the average annual profit for that city across all years? | WITH CTE AS (  SELECT \*,  EXTRACT(YEAR FROM order\_date) AS year  FROM orders  JOIN order\_details ON orders.order\_id = order\_details.order\_id  )  SELECT  year,  SUM(order\_profits) AS profits,  ROUND(AVG(order\_profits::numeric),3) AS avg\_profit  FROM CTE  WHERE shipping\_city = 'New York City'  GROUP BY year  ORDER BY year ASC  ; |
| 8 | What is the distribution of customer types in the data? | SELECT customer\_segment,COUNT(customer\_id) AS num\_customer  FROM customers  GROUP BY customer\_segment  ; |
| 9 | What’s the most profitable product category on average in Iowa across all years? | SELECT p.product\_category, AVG(od.order\_profits) avg\_profit  FROM product p  LEFT JOIN order\_details od ON p.product\_id = od.product\_id  LEFT JOIN orders o ON od.order\_id = o.order\_id  WHERE shipping\_state = 'Iowa'  GROUP BY p.product\_category  ORDER BY avg\_profit DESC  LIMIT 1; |
| 10 | What is the most popular product in that category across all states in 2016? | SELECT p.product\_id, p.product\_name, COUNT(DISTINCT o.order\_id) As Num\_orders, SUM(od.quantity) AS total\_Qty  FROM orders o  JOIN order\_details od ON o.order\_id = od.order\_id  JOIN product p ON p.product\_id = od.product\_id  WHERE EXTRACT (YEAR FROM o.order\_date)=2016 AND p.product\_category LIKE 'Furniture'  GROUP BY 1,2  ORDER BY 3 DESC, 4 DESC  ; |
| 11 | Which customer got the most discount in the data? (in total amount) | WITH CTE1 AS (  SELECT o.order\_id, o.customer\_id,  c.customer\_name, c.customer\_segment,  o.order\_date, o.shipping\_city, o.shipping\_state, o.shipping\_mode,  od.order\_details\_id, od.product\_id, od.quantity, od.order\_discount,  od.order\_profits, od.order\_profit\_ratio, od.order\_sales,  EXTRACT(YEAR FROM o.order\_date) AS year,  od.order\_sales \* od.order\_discount AS total\_discount  FROM orders o  JOIN order\_details od ON o.order\_id = od.order\_id  JOIN customers c ON o.customer\_id = c.customer\_id  )  SELECT customer\_id, customer\_name,  ROUND(SUM(total\_discount::numeric), 1) AS total\_discount,  ROUND(SUM(order\_discount::numeric), 1) AS total\_order\_discount,  ROUND(SUM(order\_sales::numeric), 1) AS total\_order\_sales,  COUNT(DISTINCT order\_id) AS total\_orders  FROM CTE1  GROUP BY customer\_id, customer\_name  ORDER BY total\_discount DESC  ; |
| 12 | How widely did monthly profits vary in 2018? | WITH CTE AS (  SELECT TO\_CHAR(order\_date,'MM-YYYY') AS month,  SUM(order\_profits) AS month\_total  FROM order\_details  JOIN orders  ON orders.order\_id = order\_details.order\_id  WHERE date\_part('year',order\_date) = '2018'  GROUP BY month  ORDER BY month )  SELECT \*, (month\_total - LAG(month\_total,1,0) OVER () ) AS month\_diff  FROM cte  -- # Alternative code. I also added early change to see how the profit ratio changed.  SELECT EXTRACT( YEAR FROM o.order\_date) AS year,  SUM(od.order\_profits) AS sum\_order\_profit,  SUM(od.order\_sales) AS total\_sales,  ROUND((SUM(od.order\_profits)/SUM(od.order\_sales))::NUMERIC,3) AS total\_profit\_ratio  FROM order\_details od  LEFT JOIN orders o ON od.order\_id = o.order\_id  GROUP BY year  ORDER BY year  ; |
| 13 | Which was the biggest order regarding sales in 2015? | SELECT o.order\_id, SUM(od.order\_sales) total\_sales  FROM order\_details od  LEFT JOIN orders o ON od.order\_id = o.order\_id  WHERE EXTRACT(YEAR FROM o.order\_date) = 2015  GROUP BY o.order\_id  ORDER BY total\_sales --DESC  LIMIT 1; |
| 14 | What was the rank of each city in the East region in 2015 in quantity? | SELECT  o.shipping\_city,  SUM(od.quantity) AS Qty,  RANK() OVER (ORDER BY SUM(quantity) DESC) AS cityrank  FROM orders o  JOIN order\_details od ON o.order\_id = od.order\_id  WHERE shipping\_region = 'East' AND EXTRACT (YEAR FROM o.order\_date)=2015  GROUP BY o.shipping\_city  ; |
| 15 | Display customer names for customers who are in the segment ‘Consumer’ or ‘Corporate.’ How many customers are there in total? | -- Ensuring unique customers with CASE WHEN (Quicker)  SELECT  COUNT(DISTINCT CASE WHEN customer\_segment = 'Consumer' THEN customer\_id END) AS Consumer\_count,  COUNT(DISTINCT CASE WHEN customer\_segment = 'Corporate' THEN customer\_id END) AS Corporate\_count,  COUNT(DISTINCT CASE WHEN customer\_segment = 'Home Office' THEN customer\_id END) AS HomeOffice\_count  FROM customers  ;  -- List of Consumers or Corporate (alphabetic order first name)  SELECT customer\_name  FROM customers  WHERE customer\_segment IN ('Consumer', 'Corporate')  GROUP BY customer\_name  ORDER BY customer\_name ASC  ; |
| 16 | Calculate the difference between the largest and smallest order quantities for product id ‘100.’ | SELECT  (MAX(quantity)-MIN(quantity)) AS qty\_difference  FROM order\_details  WHERE product\_id = 100  ; |
| 17 | Calculate the percent of products that are within the category ‘Furniture.’ | SELECT CONCAT(ROUND((SELECT COUNT(product\_category):: NUMERIC  FROM product  WHERE product\_category = 'Furniture'  GROUP BY product\_category)/COUNT(product\_category)\*100,2),'%') perc\_furniture  FROM product  ; |
| 18 | Display the number of product manufacturers with more than 1 product in the product table. | SELECT product\_manufacturer,  COUNT(DISTINCT product\_id)  FROM product  GROUP BY 1  HAVING COUNT(DISTINCT product\_id)>1  ORDER BY 2 DESC  ; |
| 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(\*) AS total\_products  FROM product  GROUP BY product\_subcategory  ORDER BY total\_products DESC, product\_subcategory ASC  ; |
| 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\_quantities  FROM order\_details  WHERE quantity >= 100  GROUP BY product\_id  ; |
| 21 | Bonus question: Join all database tables into one dataset that includes all unique columns and download it as a .csv file | WITH CTE\_ALL AS (  SELECT  o.\*, EXTRACT(YEAR FROM o.order\_date) AS year,  od.order\_details\_id, od.quantity, od.order\_discount,  od.order\_profits, od.order\_profit\_ratio, od.order\_sales,  c.customer\_name, c.customer\_segment,  p.\*  -- p.product\_id, p.product\_name, p.product\_category, p.product\_subcategory, p.product\_manufacturer  FROM orders o  JOIN order\_details od ON o.order\_id = od.order\_id  JOIN customers c ON o.customer\_id = c.customer\_id  JOIN product p ON p.product\_id = od.product\_id  )  SELECT \*  FROM CTE\_ALL  ; |