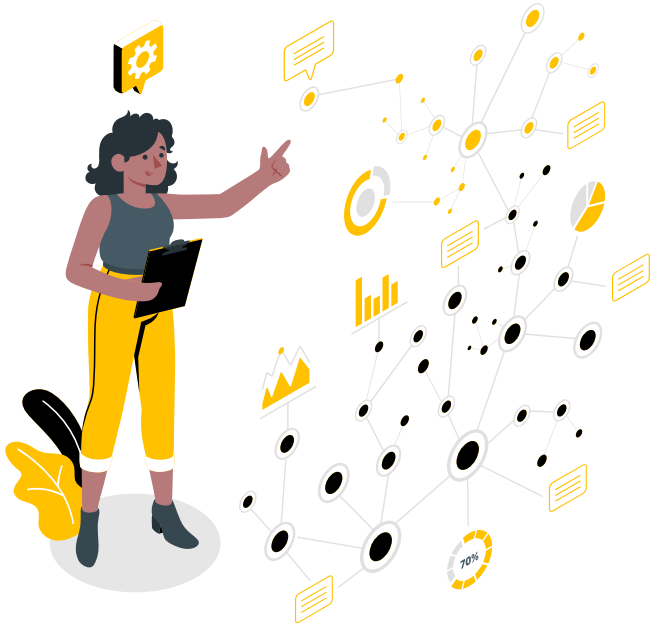
Homework SQL

AJIMAS GALUNG CANDRA

Table of Contents

Business Background

01



You can enter a subtitle here in case you need it

Business Background

BUSINESS OVERVIEW

Superstore is a global store based in USA which sells a product for human needs such as furniture, technology, and office supplies.

TASK

As a Business Intelligence Analyst, we work together with business development team, marketing, and sales to support them to analyze business needs and problem.

Business Problems

02

Problem 1

Superstore serves a SAME DAY delivery ship mode. We are asked to get the total order in SAME DAY ship mode that is delayed in delivery or ship date is more than order date (late).

Problem 2

Create a discount level, where if discount < 0.2 it is categorized as ‘LOW’, 0.2 <= discount < 0.4 is ‘MODERATE’, otherwise is ‘HIGH’ and show the relationship between discount level with average of profit received by company.

Problem 3

Working together with team sales and analyzed the average profit and average discount for each Category-Subcategory pairs.

Problem 4

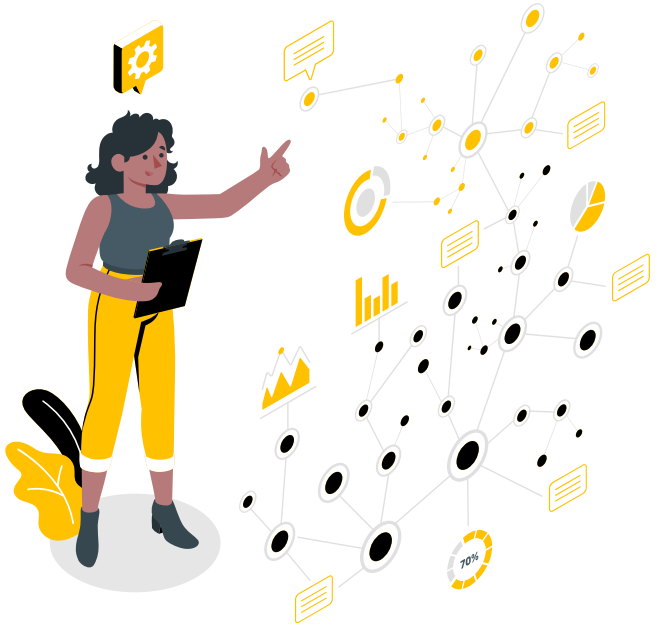
Discuss with business development team to expand the business. So, we need to analyze total sales and average profit for each customer segments in California, Texas, and Georgia in 2016.

Problem 5

Business team are interested to customer who love discount. We are asked to show total customer who has average discount more than 0.4 for each region.

Dataset Overview

03

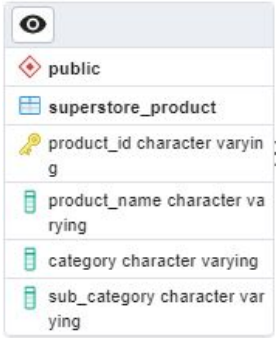


ER Diagram

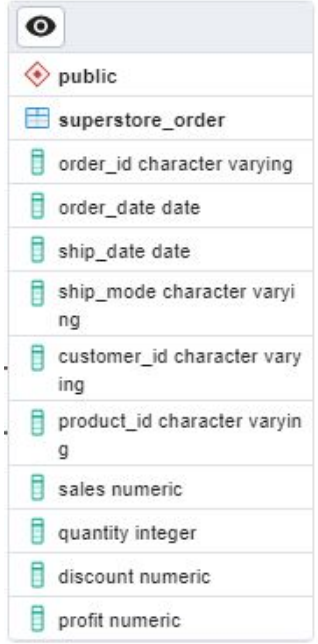

public
superstore_customer
customer_id character varying
customer_name character varying
segment character varying
country character varying
city character varying
state character varying
postalcode numeric
region character varying
public
superstore_order
order_id character varying
order_date date
ship_date date
ship_mode character varying
customer_id character varying
product_id character varying
sales numeric
quantity integer
discount numeric
profit numeric
public
superstore_product
product_id character varying
product_name character varying
category character varying
sub_category character varying

Data Descriptions

superstore_customer: it explains about the demographic of customer with the primary key is customer_id



superstore_product: it explains about the product sold by superstore with primary key is product_id



superstore_order: it explains about the orders conducted by the customer. It contains such as the product is bought, total item ordered, the order date and ship date (when the product is delivered) as well as the monetary elements (profit, sales, discount)

SQL Analysis

04

Problem 1

1. **Query**

```
SELECT COUNT(same_day.*)
FROM (
    SELECT order_id
    FROM public.superstore_order
    WHERE ship_mode = 'Same Day' AND ship_date > order_date
) AS same_day;
```

2. **Result Table**

	count bigint
1	24

Problem 2

1. **Query**

```
WITH cte_level_discount AS
(
    SELECT
        discount,
        CASE
            WHEN discount < 0.2 THEN 'LOW'
            WHEN discount >= 0.2 and discount < 0.4 THEN 'MODERATE'
            WHEN discount >= 0.4 THEN 'HIGH'
        END AS discount_level,
        profit
    FROM public.superstore_order
)
SELECT discount_level, ROUND(AVG(profit), 2) AS average_profit
FROM cte_level_discount
GROUP BY discount_level
ORDER BY average_profit DESC;
```

2. **Result**

	discount_level text	average_profit numeric
1	LOW	67.01
2	MODERATE	19.84
3	HIGH	-107.65

Problem 3

1. Query

```
SELECT
    CONCAT(p.category,'-',p.sub_category) AS category_subcategory,
    ROUND(AVG(o.discount), 3) AS average_discount,
    ROUND(AVG(o.profit), 3) AS average_profit
FROM public.superstore_order o
INNER JOIN public.superstore_product p ON o.product_id = p.product_id
GROUP BY category_subcategory
ORDER BY average_discount DESC;
```

2. Table Result

	category_subcategory text	average_discount numeric	average_profit numeric
1	Office Supplies-Binders	0.372	19.844
2	Technology-Machines	0.306	29.433
3	Furniture-Tables	0.261	-55.566
4	Furniture-Bookcases	0.211	-15.231
5	Furniture-Chairs	0.170	42.809
6	Office Supplies-Appliances	0.167	38.923
7	Technology-Copiers	0.162	817.909
8	Technology-Phones	0.155	50.074
9	Furniture-Furnishings	0.138	13.646
10	Office Supplies-Fasteners	0.082	4.376
11	Office Supplies-Envelopes	0.080	27.418
12	Technology-Accessories	0.078	54.112
13	Office Supplies-Supplies	0.077	-6.258
14	Office Supplies-Art	0.075	8.201
15	Office Supplies-Storage	0.075	25.152
16	Office Supplies-Paper	0.075	24.857
17	Office Supplies-Labels	0.069	15.237

Problem 4

1. Query

```
WITH cte_customer AS
(
    SELECT customer_id, segment
    FROM public.superstore_customer
    WHERE state IN ('California','Texas','Georgia')
)
SELECT
    c.segment AS segment,
    ROUND(SUM(o.sales), 3) AS total_sales,
    ROUND(AVG(o.profit), 2) AS average_profit
FROM public.superstore_order o
INNER JOIN cte_customer c ON o.customer_id = c.customer_id
WHERE EXTRACT(YEAR FROM o.order_date) = 2016
GROUP BY segment
ORDER BY total_sales DESC;
```

2. Table Result

	segment character varying	total_sales numeric	average_profit numeric
1	Consumer	90982.320	30.33
2	Corporate	50951.911	33.57
3	Home Office	34897.953	34.66

Problem 5

1. **Query**

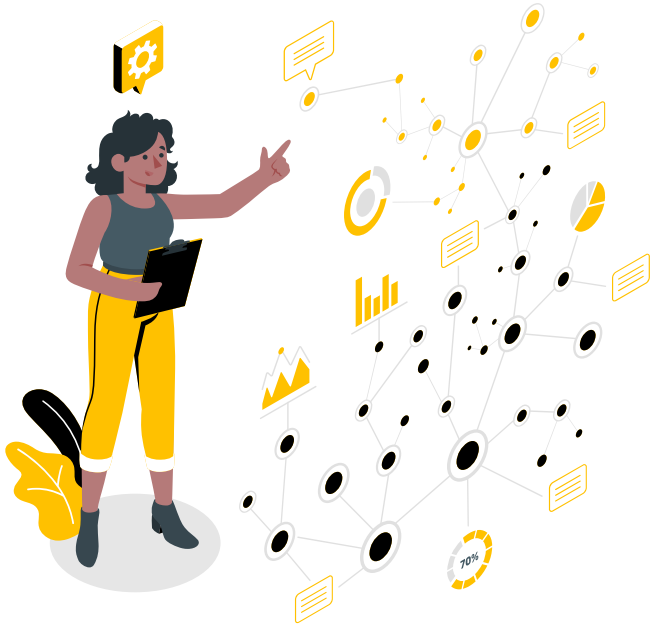
```
SELECT region, COUNT(customer_id) AS total_customer
FROM public.superstore_customer
WHERE customer_id IN (
    SELECT customer_id
    FROM public.superstore_order
    GROUP BY customer_id
    HAVING AVG(discount) >= 0.4
    )
GROUP BY region
ORDER BY total_customer DESC;
```

2. **Table Result**

	region character varying	total_customer bigint
1	West	3
2	Central	3
3	South	2
4	East	2

Result & Conclusion

05



Conclusion

1. Total orders that the ship mode is SAME DAY delivery which is delayed is **24 orders.**
2. The **higher level of discount**, **average profit decrease** and vice versa.
3. The profit for HIGH level of discount is around **-107.**
4. Categories **Technology** receive the most average profit for company with subcategory **Copiers, Accessories,** and **Phone.**
5. **HOME OFFICE** segments has the **highest average profit**. However, has the **lowest total sales** comparing to others segments.
6. **West** and **Central** has **3 customer** who the average discount is more than 0.4 and **2 customer** for **South** and **East** region.

Thank You