Sneha Parashar

**HUBBLEMIND SQL Internship**

SQL Problem Statements- Solutions

horizontal line

**Dataset Description:** This dataset contains customer data, including demographic information, purchase history, and responses to marketing campaigns. The data can be used to analyse customer behaviour, segment customers, and evaluate the effectiveness of marketing strategies.

**Dataset Link:** [**Kaggle**](https://www.kaggle.com/datasets/krsign/ecom-dataset)

**GitHub Link:**[**Solution**](https://github.com/ParasharSneha/SQL-Problem-Statement---Solutions)

**Agenda:**To use MySQL Server to solve given queries. (Source table: customer\_data)

**First created a Schema named ‘ecom’ then created a table ‘customer\_data’ with the attributes with respect to the datatype given on Kaggle.**

create database ecom;

use ecom;

show tables;

create table customer\_data (

ID int,

Year\_Birth year,

Education varchar(15),

Marital\_Status varchar(15),

Income int,

Kidhome tinyint,

Teenhome tinyint,

Dt\_Customer date,

Recency tinyint,

MntWines smallint,

MntFruits smallint,

MntMeatProducts smallint,

MntFishProducts smallint,

MntSweetProducts smallint,

MntGoldProds smallint,

NumDealsPurchases tinyint,

NumWebPurchases tinyint,

NumCatalogPurchases tinyint,

NumStorePurchases tinyint,

NumWebVisitsMonth tinyint,

AcceptedCmp3 tinyint,

AcceptedCmp4 tinyint,

AcceptedCmp5 tinyint,

AcceptedCmp1 tinyint,

AcceptedCmp2 tinyint,

Complain tinyint,

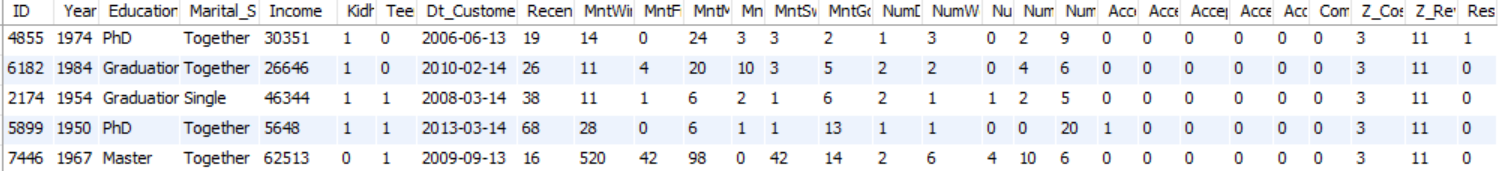
Z\_CostContact tinyint,

Z\_Revenue tinyint,

Response tinyint

);

select count(\*) from customer\_data limit 5;

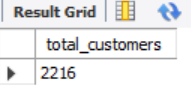
****

**Problem Statements:**

**1. Count the number of customers:**

**Task:** Count the total number of customers in the dataset.

**Query:** **SELECT COUNT**(ID) **AS** total\_customers **FROM** customer\_data;



**Explanation:**

* **Purpose**: Above query counts the total number of rows in the **customer\_data**

table. Since each row represents a unique customer, the output will give the

total number of customers in the dataset.

* **Why It's Used**: Knowing the total number of customers helps in understanding

the dataset's size and provides context for further analysis.

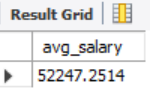
* **Expected Result**: A single integer representing the total number of customers (Here total customers are 2216).

----------------------------------------------------------------------------------------------------------------------------

**2. Find the average income:**

**Task:** Calculate the average income of all customers.

**Query: SELECT AVG**(Income) **AS** avg\_salary **FROM** customer\_data;



**Explanation:** To calculate the average of any numeric value we have AVG() function.

* **Purpose**: Above query calculates the average income of all customers.
* **Why It's Used**:Knowing the average income helps in understanding the

economic profile of the customers, which is useful for tailoring marketing

strategies.

* **Expected Result**: A single value representing the average income of the

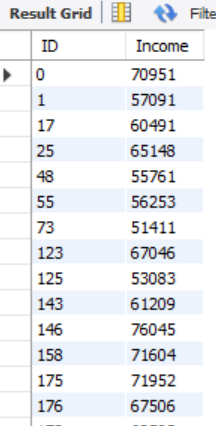
customers.

----------------------------------------------------------------------------------------------------------------------------

**3. List customers with income above $50,000:**

**Task:** Retrieve the IDs of customers whose income is above $50,000.

**Query: SELECT** ID, Income **FROM** customer\_data **WHERE** income > 50000;



**Explanation:** To filter the data where use the WHERE clause(Here I got 1156 records).

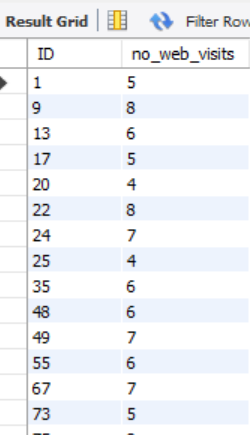
* **Purpose**: Above query retrieves customer IDs and their Income above $50000.
* **Why It's Used**: Identifying high-income customers can help the marketing team to target specific customers.
* **Expected Result**: A list of customer IDs and their Income.

----------------------------------------------------------------------------------------------------------------------------

**4. Find customers with more than 3 web visits per month:**

**Task:** Identify customers who visit the website more than three times a Month.

**Query: SELECT** ID, NumWebVisitsMonth **AS** no\_web\_visits **FROM** customer\_data **WHERE** NumWebVisitsMonth > 3 **ORDER BY** ID;



**Explanation:** Here ,I applied the Where clause because the NumWebVisitsMonthcolumn stands for Number of visits to a company's website in the last month. After applying this query I got 1652 records.

* **Purpose**: Above query retrieves customer IDs and number of times they visit the web more than 3 visits per month.
* **Why It's Used**: Frequent web visitors are likely more engaged with the

Company’s offering, making them key targets for personalised marketing efforts.

* **Expected Result**: A list of customer IDs with their corresponding number of web visits.

----------------------------------------------------------------------------------------------------------------------------

**5. List customers who have accepted at least one campaign:**

**Task:** Retrieve customers who have accepted at least one marketing campaign.

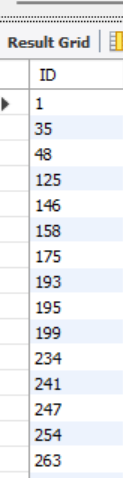
**Query1: SELECT** ID **FROM** customer\_data **WHERE** AcceptedCmp1 = 1 **OR**

AcceptedCmp2 = 1 **OR**  AcceptedCmp3 = 1 **OR** AcceptedCmp4 = 1 **OR**

AcceptedCmp5 = 1 **ORDER BY** ID;

**Query2: SELECT** ID **FROM** customer\_data **WHERE** 1 **IN** (AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4, AcceptedCmp5)

**ORDER BY** ID**;**



**Explanation:** Here, I use OR logical operator and IN operator to meet the condition that either a customer takes part in at least 1 campaign (Here I got 459 records).

* **Purpose**: Above query retrieves the customer IDs who have accepted at least 1 campaign.
* **Why It's Used**:Identifying customers who respond positively to campaigns can

help in refining future marketing efforts and targeting.

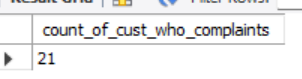
* **Expected Result**: A list of customer IDs who have accepted at least 1 campaign.

----------------------------------------------------------------------------------------------------------------------------

**6. Count customers with complaints:**

**Task:** Count the number of customers who have made complaints.

**Query: SELECT COUNT**(ID) **AS** count\_of\_cust\_who\_complaints **FROM** customer\_data **WHERE** Complain = 1;

****

**Explanation:**

* **Purpose**: This query counts the number of customers who have made

complaints by filtering the dataset where the Complain column has a value of 1.

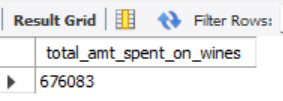
* **Why It's Used**: Understanding the number of customers who have raised complaints helps in evaluating customer satisfaction and areas that might need improvement.
* **Expected Result**: A single integer representing the total number of customers who have made complaints

----------------------------------------------------------------------------------------------------------------------------

**7. Calculate total amount spent on wines:**

**Task:** Calculate the total amount spent on wines by all customers.

**Query:** **SELECT SUM**(MntWines) **AS** total\_amt\_spent\_on\_wines **FROM** customer\_data;

****

**Explanation:** To calculate the total amount spent on wines I use the SUM function.

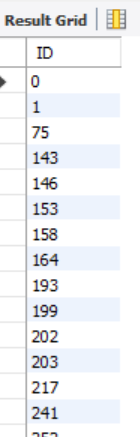
* **Purpose**: Above query calculates the total amount spent on wines.
* **Why It's Used**: Understanding the total expenditure on specific product categories helps in assessing customer preferences and sales performance.
* **Expected Result**: A single value representing the total amount spent on wines.

----------------------------------------------------------------------------------------------------------------------------

**8. Find customers with no children at home:**

**Task:** List customers who do not have any kids or teenagers at home.

**Query: SELECT** ID **FROM** customer\_data **WHERE** Kidhome=0 **AND** Teenhome=0 **ORDER BY** ID;

****

**Explanation:** There are 633 customers who do not have kids or teenagers at home.

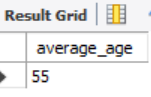
* **Purpose**: Above query retrieves the IDs of customers who have no children at home.
* **Why It's Used**: Knowing which customers have no dependents at home might influence marketing strategies, as these customers may have different purchasing behaviours.
* **Expected Result**: A list of customer IDs who have no children or teenagers at home.

----------------------------------------------------------------------------------------------------------------------------

**9. Determine the average age of customers:**

**Task:** Calculate the average age of customers.

**Query: SELECT ROUND((AVG**((**YEAR**(**NOW()**) - Year\_Birth)),0) **AS** average\_age **FROM** customer\_data;

****

**Explanation:** In SQL we have AVG function to calculate average and NOW() to find current date and YEAR function to extract year from the date.

* **Purpose**: This query calculates the average age of customers by subtracting the Year\_Birth from the current year (2024) and then averaging the results.
* **Why It's Used**: Knowing the average age of customers helps in understanding the demographic profile, which is crucial for tailoring products and services.
* **Expected Result**: A single value representing the average age of customers.

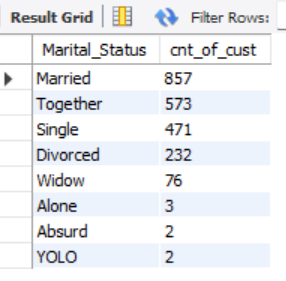
----------------------------------------------------------------------------------------------------------------------------

**10. List customers by marital status:**

**Task:** Group customers by their marital status and count the number of

customers in each group.

**Query: SELECT** Marital\_Status, **COUNT**(ID) **AS** cnt\_of\_cust **FROM** customer\_data **GROUP BY** Marital\_Status **ORDER BY COUNT**(ID) **DESC**;



**Explanation:** Here I use GROUP BY clause to group the data as per their marital status.

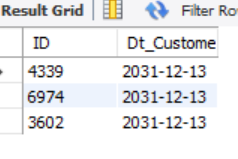
* **Purpose**: Above query groups customers by their marital status and counts the number of customers in each group.
* **Why It's Used**: Understanding the distribution of customers by marital status can help in segmenting the market and developing targeted marketing strategies.
* **Expected Result**: A list showing the number of customers in each marital status category.

----------------------------------------------------------------------------------------------------------------------------

**11. Find the most recent customer:**

**Task:** Identify the customer who joined most recently.

**Query: SELECT** ID, Dt\_Customer **FROM** customer\_data **WHERE** Dt\_Customer= (**SELECT** **MAX**(Dt\_Customer) **FROM** customer\_data);



**Explanation:** Here I use Subquery because there are multiple customers who join on the same date (Here i got 3 most recent customers who join on the same date).

* **Purpose**: Above query identifies the most recent customer/customers as per date of joining.
* **Why It's Used**: Knowing the most recent customer can be useful for understanding recent trends in customer acquisition.
* **Expected Result**: ID and date of joining of the most recent customer or list or customers.

----------------------------------------------------------------------------------------------------------------------------

**12. Calculate the total amount spent on each product category:**

**Task:** Calculate the total amount spent on each product category (Wines,

Fruits, Meat, Fish, Sweets, Gold).

**Query: SELECT**  **SUM**(MntWines) **AS** total\_spent\_on\_wines,

**SUM**(MntFruits) **AS** total\_spent\_on\_fruits,

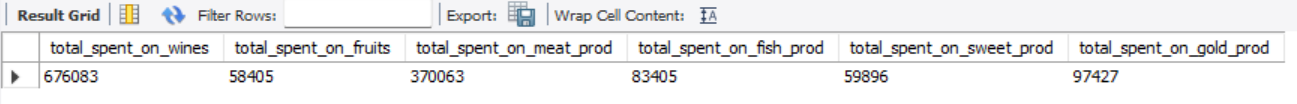
**SUM**(MntMeatProducts) **AS** total\_spent\_on\_meat\_prod,

**SUM**(MntFishProducts) **AS** total\_spent\_on\_fish\_prod,

**SUM**(MntSweetProducts) **AS** total\_spent\_on\_sweet\_prod,

**SUM**(MntGoldProds) **AS** total\_spent\_on\_gold\_prod

**FROM** customer\_data;



**Explanation:** Here to calculate spent on each category I just apply SUM function to all the necessary columns.

* **Purpose**: Above query calculates the total amount spent on each product.
* **Why It's Used**: Knowing how much customers spend on different product categories allows the business to understand consumer preferences. This insight can help in stock management and marketing strategies.
* **Expected Result**: Total expenditure in each product category.

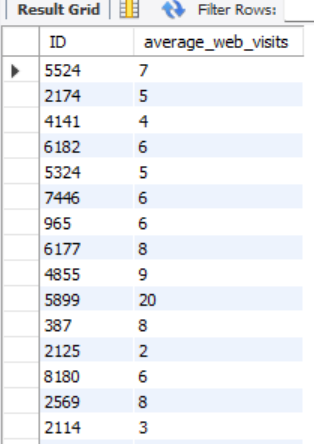
----------------------------------------------------------------------------------------------------------------------------

**13. Find the average number of web visits for each customer:**

**Task:** Calculate the average number of web visits per month for each

customer.

**Query: SELECT** ID, **ROUND**(**AVG**(NumWebVisitsMonth),0) **AS** average\_web\_visits **FROM** customer\_data **GROUP BY** ID;

****

**Explanation:**  above I’ve used AVG() to calculate the average and then round off the result using ROUND() function. Here I get 2216 records.

* **Purpose**: Above query calculates the average number of web visits by each customer by grouping the dataset by ID.
* **Why It's Used**: Understanding the average web visits helps in identifying customer engagement levels. This information is useful for targeting customers who are more active online.
* **Expected Result**: A list of customer IDs along with their average number of web visits last month.

----------------------------------------------------------------------------------------------------------------------------

**14. Identify high-value customers:**

**Task:** List customers who have spent more than $2,000 in total.

**Query: WITH** TotalAmount **AS** (

**SELECT** ID ,

(MntWines + MntFruits + MntMeatProducts + MntFishProducts

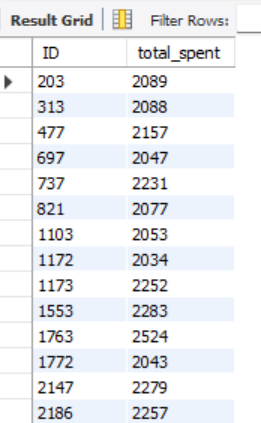
+ MntSweetProducts + MntGoldProds ) **AS** total\_spent

**FROM** customer\_data **ORDER BY** ID

)

**SELECT** \* **FROM** TotalAmount

**WHERE** total\_spent > 2000;



**Explanation:** Here I use CTE to find the total spent by each customer then apply the given condition on the total spent (Here I get 50 records).

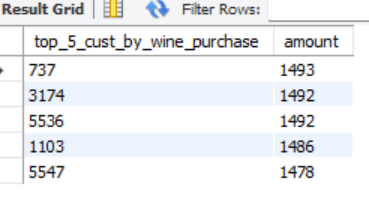
* **Purpose**: Above query identifies high-value customers by calculating their total spending across all product categories who have spent more than $2,000.
* **Why It's Used**: High-value customers contribute significantly to revenue, and identifying them allows the business to focus on retention and personalised marketing for these valuable clients.
* **Expected Result**: A list of customer IDs and their corresponding spends on all products.

----------------------------------------------------------------------------------------------------------------------------

**15. List top 5 customers by wine purchases:**

**Task:** Identify the top 5 customers who spent the most on wines.

**Query: SELECT** ID **AS** top\_5\_cust\_by\_wine\_purchase, MntWines **AS** amount **FROM** customer\_data **ORDER BY** MntWines **DESC LIMIT** 5;



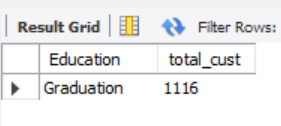
**Explanation:** Here uses the ORDER BY clause in Descending order(DESC) on the column with respect to which we want to sort the data and as we need to find only top 5 customers so here I have set the LIMIT to 5.

* **Purpose**: Above query retrieves the top 5 customers who spend most on wines.
* **Why It's Used**:Identifying the top spenders in a specific product category, such as wines, allows the business to target these customers with special offers or loyalty programs.
* **Expected Result**: A list of customer IDs and their total spent on wine products.

----------------------------------------------------------------------------------------------------------------------------

**16. Find the most common education level:Task:** Determine the most common education level among customers.

**Query: SELECT** Education, **COUNT**(ID) **AS** total\_cust **FROM** customer\_data **GROUP BY** Education **ORDER BY** COUNT(ID) DESC **LIMIT** 1;



**Explanation:** Here we need to find the most common education level which means the level which has the maximum number of customers. So I use the Group By clause and limit to 1.

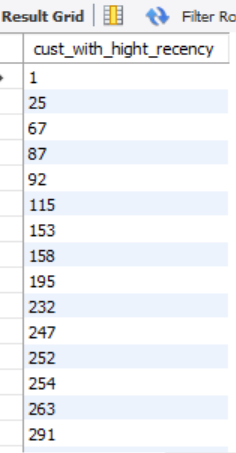
* **Purpose**: Above query determines the most common education level among customers.
* **Why It's Used**: Understanding the educational background of customers helps in creating tailored marketing messages and products that align with the educational profile of the target audience.
* **Expected Result**: The most common education level and the number of customers with that education.

----------------------------------------------------------------------------------------------------------------------------

**17. Identify customers with high recency:**

**Task:** List customers who have recently interacted with the company (Recency < 30 days).

**Query: SELECT** ID **AS** cust\_with\_hight\_recency **FROM** customer\_data **WHERE** Recency < 30 **ORDER BY** ID;

**......**

**Explanation:** Here we have the condition on the Recency column. Here I got a total of 686 records.

* **Purpose**: Above query identifies the customers who have interacted with the company recently (within 30 days).
* **Why It's Used**:Recent interactions are a strong indicator of customer engagement. These customers are prime targets for follow-up marketing efforts to maintain their interest.
* **Expected Result**: A list of customer IDs who integrated with the company within the last 30 days.

----------------------------------------------------------------------------------------------------------------------------

**18. Calculate the average number of accepted campaigns per customer:**

**Task:** Determine the average number of campaigns accepted by customers.

**Query: WITH** TotalCamp **AS** (

**SELECT** ID,

(

**COALESCE**(AcceptedCmp1,0) +

**COALESCE**(AcceptedCmp2,0) +

**COALESCE**(AcceptedCmp3,0) +

**COALESCE**(AcceptedCmp4,0) +

**COALESCE**(AcceptedCmp5,0)

) **AS** cnt\_camp

**FROM** customer\_data

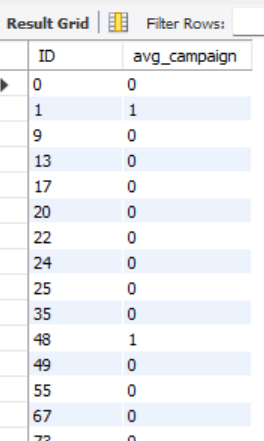
)

**SELECT** ID, **CEIL**(**AVG**(cnt\_camp)) **AS** avg\_campaign

**FROM** TotalCamp

**GROUP BY** ID

**ORDER BY** ID;

.....

**Explanation:** Here I use CTE to calculate the total number of campaigns accepted by customers and use COALESCE to handle the Null values.

* **Purpose**: Above query calculates the average number of campaigns accepted by each customer.
* **Why It's Used**:Understanding how many campaigns customers typically accept helps in evaluating the effectiveness of marketing campaigns and customer responsiveness
* **Expected Result**: A list of customer IDs with average number of campaign accepted by them.

----------------------------------------------------------------------------------------------------------------------------

**19. Find customers with the highest total purchases:**

**Task:** Identify customers with the highest total number of purchases.

**Query: WITH** TotalPurchase **AS** (

**SELECT** ID,

(

**COALESCE**(NumDealsPurchases,0) +

**COALESCE**(NumWebPurchases,0) +

**COALESCE**(NumCatalogPurchases,0) +

**COALESCE**(NumStorePurchases,0)

) **AS** ttl\_pur

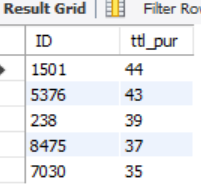
**FROM** customer\_data

)

**SELECT** \* **FROM** TotalPurchase

**ORDER BY** ttl\_pur **DESC**

**LIMIT** 5;

****

**Explanation:** Here are the Top 5 customers who made the highest purchase.

* **Purpose**: Above query identifies customers with the highest number of purchases

across all product categories.

* **Why It's Used**: Knowing the top purchasers allows the business to focus retention strategies on these valuable customers, potentially offering them exclusive deals or rewards.
* **Expected Result**: A list of the top 5 customers with their total number of purchases.

----------------------------------------------------------------------------------------------------------------------------

**20. List customers by their response to the last campaign:**

**Task:** Group customers based on their response to the last campaign.

**Query:** **SELECT CASE**

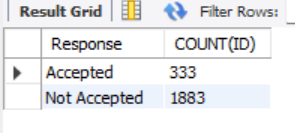
**WHEN** Response=1 **THEN** 'Accepted'

**ELSE** 'Not Accepted'

**END AS** Response

, **COUNT**(ID)

**FROM** customer\_data **GROUP BY** Response;



**Explanation:** Here I’ve used Case to rename the response 1 to Accepted and 0 to Not Accepted.

* **Purpose**: Above query groups customers based on their response to the last marketing campaign and counts the number of customers in each group.
* **Why It's Used**: Understanding how customers responded to the most recent

campaign helps in assessing its effectiveness and refining future marketing strategies.

* **Expected Result**: A breakdown of customer responses to the last campaign.

****