

Виконання

представлення у схемі Students:

```
SELECT * FROM `data-analytics-mate.Students.v_Bozhko_SQL Advanced Module Task`
```

створення представлення

```
CREATE VIEW `Students.v_Bozhko_SQL Advanced Module Task`  
as  
"основний код SQL-запиту"
```

основний код SQL-запиту :

```
WITH account AS  
-- інформація по акаунтах  
--CTE account У цьому етапі ми об'єднуємо дані з таблиць DA.account,  
DA.account_session, DA.session та DA.session_params, щоб отримати інформацію про  
акаунти за датою, країною, інтервалом розсилки та статусами верифікації та  
відписки. Вихідні дані групуються за комбінаціями полів s.date, sp.country,  
ac.send_interval, ac.is_verified та ac.is_unsubscribed, і підраховується кількість  
унікальних акаунтів (account_cnt).  
(  
  SELECT  
    s.date,  
    sp.country,  
    COUNT(DISTINCT a.id) AS account_cnt,  
    a.send_interval AS send_interval,  
    a.is_verified AS is_verified,  
    a.is_unsubscribed AS is_unsubscribed  
  FROM `data-analytics-mate.DA.account` a  
  JOIN `DA.account_session` acs  
  on a.id = acs.account_id  
  join `DA.session` s  
  on acs.ga_session_id = s.ga_session_id  
  join `DA.session_params` sp  
  ON acs.ga_session_id = sp.ga_session_id  
  GROUP BY s.date, sp.country, send_interval, is_verified, is_unsubscribed  
)  
  
message AS
```

--інформація по листах

--CTE message (email_metrics) На цьому етапі збирається інформація про активність електронної пошти (кількість надісланих, відкритих та переглянутих повідомлень) за країною, датою, інтервалом розсилки та іншими параметрами. Ми об'єднуємо таблиці DA.email_sent, DA.email_open, DA.email_visit, DA.account_session, DA.session, DA.session_params та DA.account, щоб сформувати ці метрики. Результат групується за комбінацією полів, таких як sent_date, sp.country, ac.send_interval, ac.is_verified та ac.is_unsubscribed.

```
(
    SELECT
        DATE_ADD(s.date, INTERVAL ems.sent_date DAY) AS sent_date,
        sp.country,
        ems.sent_date as send_interval,
        a.is_verified AS is_verified,
        a.is_unsubscribed AS is_unsubscribed,
        COUNT(DISTINCT ems.id_message) AS sent_msg,
        COUNT(DISTINCT eo.id_message) AS open_msg,
        COUNT(DISTINCT ev.id_message) AS visit_msg

    FROM    `data-analytics-mate.DA.account` a
    JOIN    `DA.account_session` acs
        on a.id = acs.account_id
    join `DA.session_params` sp
    ON acs.ga_session_id = sp.ga_session_id
    JOIN    `DA.session` s
    ON     acs.ga_session_id = s.ga_session_id
    join `DA.email_sent` ems
    ON     ems.id_account = acs.account_id
    LEFT JOIN    `DA.email_open` eo
    ON     ems.id_message = eo.id_message
    LEFT JOIN    `DA.email_visit` ev
    ON     ems.id_message = ev.id_message
    GROUP BY    sent_date, sp.country, send_interval, is_verified, is_unsubscribed
),
```

unions as

--дві частини (по акаунтах і листах) поєднані UNION

--CTE unions Цей етап об'єднує результати з попередніх етапів account та email_metrics. Вихідні дані об'єднуються за допомогою оператора UNION ALL, щоб отримати повну таблицю з усіма метриками та параметрами.

```
(
    SELECT
        account.date AS date,
        account.country as country,
        account.account_cnt,
        account.send_interval AS send_interval,
        account.is_verified AS is_verified,
        account.is_unsubscribed AS is_unsubscribed,
```

```

    0 AS sent_msg,
    0 AS open_msg,
    0 AS visit_msg
FROM account
UNION ALL
SELECT
    sent_date AS date,
    message.country AS country,
    0 AS account_cnt,
    message.send_interval AS send_interval,
    message.is_verified AS is_verified,
    message.is_unsubscribed AS is_unsubscribed,

    message.sent_msg,
    message.open_msg,
    message.visit_msg
FROM message
),

```

final_groups as

--загальна кількість створених акаунтів та відправлених листів по країні;
 --CTE final_groups У цьому етапі ми підсумовуємо об'єднані дані за датою, країною, інтервалом розсилки та іншими параметрами. Результатом є зведена таблиця, в якій обчислюються загальні кількості для кожної комбінації параметрів.

```

(
SELECT
    date,
    country,
    send_interval,
    is_verified,
    is_unsubscribed,
    SUM(account_cnt) AS account_cnt,
    SUM(sent_msg) AS sent_msg,
    SUM(open_msg) AS open_msg,
    SUM(visit_msg) AS visit_msg
FROM unions
GROUP BY date,
    country,
    send_interval,
    is_verified,
    is_unsubscribed
),

```

sums AS

-- використання віконний функцій для ранжування країн
 -- CTE sums Цей етап обчислює загальну кількість акаунтів (total_country_account_cnt) і надісланих повідомлень (total_country_sent_cnt) для

кожної країни, використовуючи аналітичні функції SUM OVER. Після цього застосовується DENSE_RANK, щоб кожній країні присвоїти рейтинг за кількістю акаунтів і кількістю надісланих повідомлень.

```
(  
  SELECT  
    date,  
    country,  
    send_interval,  
    is_verified,  
    is_unsubscribed,  
    account_cnt,  
    sent_msg,  
    open_msg,  
    visit_msg,  
    SUM(account_cnt) OVER(PARTITION BY country ) AS total_country_account_cnt,  
    SUM(sent_msg) OVER(PARTITION BY country ) AS total_country_sent_cnt,  
    DENSE_RANK() OVER( ORDER BY account_cnt DESC) AS rank_total_country_account_cnt,  
    DENSE_RANK() OVER( ORDER BY sent_msg DESC) AS rank_total_country_sent_cnt  
  FROM final_groups  
)
```

--Фінальний запит Фінальна частина запиту вибирає рядки з таблиці sums, де ранг країни за кількістю акаунтів або за кількістю надісланих повідомлень входить у топ-10. Таким чином, запит формує результати з показниками активності електронної пошти та акаунтів для аналізу за країною, відображаючи найбільш активні країни за кількістю акаунтів або за кількістю надісланих електронних листів.

```
SELECT *  
FROM sums  
where rank_total_country_account_cnt <= 10 or rank_total_country_sent_cnt <=10
```

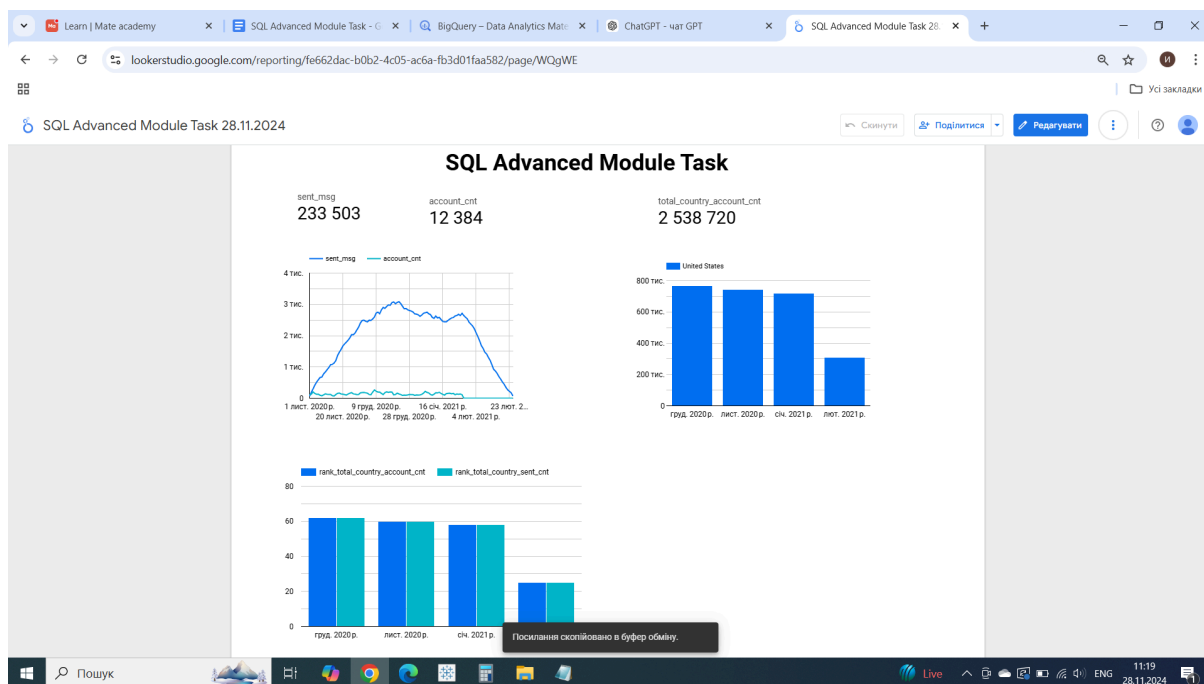
скріншот результату:

The screenshot shows the Google Cloud BigQuery console. The query results are displayed in a table with the following columns: Row, date, country, send_interval, is_verified, is_unsubscribed, account_cnt, sent_msg, open_msg, and visit_msg. The data is filtered for the United States and sorted by account_cnt in descending order.

Row	date	country	send_interval	is_verified	is_unsubscribed	account_cnt	sent_msg	open_msg	visit_msg
1	2020-12-15	United States	1	1	0	114	88	48	
2	2021-01-06	United States	1	1	0	116	95	45	
3	2020-12-13	United States	5	1	0	0	172	77	
4	2020-12-30	United States	21	1	0	0	162	58	
5	2020-12-22	United States	13	1	0	0	166	66	
6	2020-12-07	United States	1	1	0	105	88	38	
7	2020-12-15	United States	7	1	0	1	167	75	
8	2021-01-14	United States	1	1	0	113	121	71	
9	2020-12-21	United States	12	1	0	0	162	64	
10	2020-12-08	United States	1	1	0	133	93	44	

Скріншот створеної візуалізації в Looker Studio.

<https://lookerstudio.google.com/reporting/fe662dac-b0b2-4c05-ac6a-fb3d01faa582>



Завдання модуля SQL Advanced

Для виконання завдання використовуй е-commerce базу даних, з якою ми працювали в BigQuery. Напиши один SQL-запит для створення набору даних. Додатково створи візуалізацію в Looker Studio на основі отриманих даних.

Опис

Необхідно зібрати дані, які допоможуть аналізувати динаміку створення акаунтів, активність користувачів за листами (відправлення, відкриття, переходи), а також оцінювати поведінку в категоріях, таких як інтервал відправлення, верифікація акаунтів і статус підписки. Дані дозволять порівнювати активність між країнами, визначати ключові ринки, сегментувати користувачів за різними параметрами.

Вимоги до SQL-запиту

Результат запиту повинен містити певний перелік полів групування - тобто категоріальних значень (з якими не проводиться ніяких обчислень). Основні метрики по акаунтам, а також основні метрики по емейлам, повинні розраховуватися в цих розрізах:

date — дата (для акаунтів - дата створення акаунта, для емейлів - дата відправки емейла);
country — країна;
send_interval — інтервал відправлення, встановлений акаунтом;
is_verified — перевірено акаунт чи ні;
is_unsubscribed — підписник відписався чи ні.

Важливо! Ці розрізи стосуються як метрик акаунтів, так і метрик емейлів. Потрібна інформація є в таблиці `account`.

Твій запит має розраховувати інформацію (в розрізі перелічених полів) по наступним основним метриках:

account_cnt — кількість створених акаунтів;
sent_msg — кількість відправлених листів;
open_msg — кількість відкритих листів;
visit_msg — кількість переходів по листах;

А також по додатковим метриках (які розраховуються на базі основних метрик):

total_country_account_cnt — загальна кількість створених підписників в цілому по країні;
total_country_sent_cnt — загальна кількість відправлених листів в цілому по країні;
rank_total_country_account_cnt — рейтинг країн за кількістю створених підписників в цілому по країні;
rank_total_country_sent_cnt — рейтинг країн за кількістю відправлених листів в цілому по країні.

Метрики для акаунтів і емейлів потрібно розраховувати окремо, щоб зберегти унікальні розрізи для кожного та уникнути конфліктів через різну логіку використання поля date. Для об'єднання результатів використовуй `UNION`. У фіналі залиш лише ті записи, де `rank_total_country_account_cnt` або `rank_total_country_sent_cnt` менше або дорівнює 10.

Обов'язково використай хоча б одне СТЕ, винеси туди логічні частини запиту. Для розрахунку рангу використовуй функції-вікна.

Структура набору

Запит повинен виводити такі стовпчики:

date — дата;
country — країна;
send_interval — інтервал відправлення;
is_verified — перевірено акаунт чи ні;
is_unsubscribed — підписник відписався;
account_cnt — кількість створених акаунтів;
sent_msg — кількість відправлених листів;
open_msg — кількість відкритих листів;
visit_msg — кількість переходів по листах;
total_country_account_cnt — загальна кількість створених підписників по країні;
total_country_sent_cnt — загальна кількість відправлених листів по країні;
rank_total_country_account_cnt — рейтинг країн за кількістю створених підписників;
rank_total_country_sent_cnt — рейтинг країн за кількістю відправлених листів.

Візуалізація

Створи візуалізацію в Looker Studio, яка показуватиме загальні значення в розрізі країн для таких полів:

account_cnt;
total_country_sent_cnt;
rank_total_country_account_cnt;
rank_total_country_sent_cnt.

Покажи динаміку по даті для поля sent_msg.

Формат оформлення результатів

1. **Документ:** додай код SQL-запиту в документ. Додай коментарі до основних частин коду для пояснення логіки.
2. **Скріншот:** додай в документ скріншот створеної візуалізації в Looker Studio.

! Посилання на документ з описом результатів прикріпи в завданні SQL Advanced Module Task.

