

# Google Big Query

Hi everyone,

My name is Adiba Kamalia and I'm a data analyst working on an EcommerceU project by RevoU. I'm excited to be a part of this project. I'm confident that I can bring a great deal of knowledge and insight in this project and look forward to collaborating with everyone. I will continue to practice and learn in order to improve my skills in data analysis, particularly SQL.

Thanks!

## Table name

### Table Event

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">event_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">event_name</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">timestamp_event</a>	TIMESTAMP	NULLABLE
<input type="checkbox"/>	<a href="#">user_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">favorite_flag</a>	STRING	NULLABLE

### Table Transaction

Write the anticipated big-picture goals or outcomes from the project

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">event_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">event_name</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">timestamp_event</a>	TIMESTAMP	NULLABLE
<input type="checkbox"/>	<a href="#">user_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">payment_method_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">transaction_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">voucher</a>	BOOLEAN	NULLABLE
<input type="checkbox"/>	<a href="#">voucher_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">voucher_amount</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">original_amount</a>	INTEGER	NULLABLE

<input type="checkbox"/>	<a href="#">paid_amount</a>	INTEGER	NULLABLE
--------------------------	-----------------------------	---------	----------

## Table merchant

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">merchant_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">merchant_name</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">province</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">city</a>	STRING	NULLABLE

## Table Transaction\_Item

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">transaction_item_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">transaction_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">event_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">user_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">event_name</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">timestamp_event</a>	TIMESTAMP	NULLABLE
<input type="checkbox"/>	<a href="#">payment_method_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">product_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">qty</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">price</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">original_amount</a>	INTEGER	NULLABLE

## Table User

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">user_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">register_date</a>	DATE	NULLABLE
<input type="checkbox"/>	<a href="#">name</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">gender</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">province</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">city</a>	STRING	NULLABLE

## Table Product

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">product_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">product</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">product_category</a>	INTEGER	NULLABLE

<input type="checkbox"/>	<a href="#">price</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">merchant_id</a>	INTEGER	NULLABLE

## Table Product\_Category

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">product_category_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">product_category</a>	STRING	NULLABLE

## Table Payment\_method

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">payment_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">payment_method</a>	STRING	NULLABLE

## Table Voucher\_type

<input type="checkbox"/>	Field name	Type	Mode
<input type="checkbox"/>	<a href="#">voucher_id</a>	INTEGER	NULLABLE
<input type="checkbox"/>	<a href="#">voucher_type</a>	STRING	NULLABLE
<input type="checkbox"/>	<a href="#">voucher_amount</a>	INTEGER	NULLABLE

Data Time\_frame : 2019-2021

## SQL Query

Write a query to find the total transactions for each payment method in 2021, sort from the highest

### Query

```

SELECT
B.payment_method as payment_method,
count(distinct A.transaction_id) as total_transaction,
FROM `revoufsda-ecommerceu.EcommerceU_B3.transaction` as A
join `revoufsda-ecommerceu.EcommerceU_B3.payment_method` as B
on A.payment_method_id = B.payment_id
where extract(year from A.timestamp_event) in (2021)
group by 1
order by 2 DESC;

```

## Result

payment_method	total_transaction
transfer	1960
paylater	1521
e-wallet	1084
credit card	909
debit card	93

Write a query to find monthly TOP 3 Users based on total GMV.

## Query

```
with monthly as
(select*
from (
  select
format_timestamp('%Y%m',A.timestamp_event) as
year_month,
B.name as User,
sum(A.qty)*sum(A.price) as GMV,
FROM
`revoufsda-ecommerceeu.EcommerceU_B3.transaction_items` as A
join `revoufsda-ecommerceeu.EcommerceU_B3.user` as B
on A.user_id = B.user_id
where extract(year from A.timestamp_event) in
(2019,2020,2021)
group by 1,2
))

,rank_3 as
(SELECT *,
RANK() OVER(Partition by year_month ORDER BY GMV desc)
AS rank_per_month
from monthly
)
select *
from rank_3
where rank_per_month <4
order by year_month
```

## Result

year_month	User	GMV	rank_per_m...
201901	Hartaka Uwais	8610000	1
201901	Jarwa Permadji	7800000	2
201901	Ulva Yuniar S.Gz	7020000	3
201902	Cager Latupono	8100000	1
201902	Pangestu Hari Wijaya	7310000	2
201902	Indra Pradana	6900000	3
201903	Hartaka Saragih	8100000	1
201903	Warji Harimurti Utama	7030000	2
201903	Jarwadi Dabukke	6800000	3
201904	Estiono Saragih	14030000	1
201904	Karsa Mansur	13350000	2
201904	Mustika Marbun S.Kom	13230000	3
201905	Rafid Puriwanto Surveono S.Pd	23115000	1

201905	RANDI FERWANDA SURYONO,S.P.T.	201905	1
201905	Bajragin Ozy Jailani	22860000	2
201905	Caturanona Hakim S.Ked	14700000	3

## Write a Query to find monthly active user and it's growth in 2021

### Query

```

WITH
monthly_MAU AS (
SELECT
format_timestamp('%Y%m',timestamp_event) as month,
COUNT(DISTINCT user_id) as MAU
FROM `revoufsda-ecommerceeu.EcommerceU_B3.event`
where timestamp_event between '2021-01-01' AND '2021-12-31'
group by 1
order by 2 desc)
SELECT
*,(
ROUND((MAU - LAG(MAU) OVER(ORDER BY month ASC)) / LAG(MAU)
OVER(ORDER BY month ASC) * 100,2)|| '%' AS growth_rate
FROM
monthly_MAU
ORDER BY
month ASC;

```

### Result

Row	month	MAU	growth_rate
1	202101	680	null
2	202102	678	-0.29%
3	202103	674	-0.59%
4	202104	700	3.86%
5	202105	721	3%
6	202106	696	-3.47%
7	202107	842	20.98%
8	202108	812	-3.56%
9	202109	838	3.2%
10	202110	847	1.07%
11	202111	856	1.06%
12	202112	862	0.7%

## Write a Query to find AOV and ABS in 2021

### Query

```

select
format_timestamp('%Y%m',A.timestamp_event) as Month,
round (sum(B.qty)/count(B.transaction_id),2) as ABS,
round (sum(A.paid_amount)/count(A.transaction_id),2) as AOV
from `revoufsda-ecommerceeu.EcommerceU_B3.transaction` as A
join `revoufsda-ecommerceeu.EcommerceU_B3.transaction_items` as B
on A.transaction_id = B.transaction_id
where A.timestamp_event between '2021-01-01' AND '2021-12-31'
and A.event_name = 'complete'
group by 1
order by 1;

```

## Result

Row	Month	ABS	AOV
1	202101	2.84	726107.93
2	202102	2.68	699096.94
3	202103	2.83	752384.11
4	202104	2.97	759104.47
5	202105	2.91	752952.75
6	202106	2.88	762524.3
7	202107	2.79	707077.99
8	202108	2.74	673411.49
9	202109	2.94	786434.17
10	202110	2.73	710943.75
11	202111	2.82	670507.94
12	202112	2.87	746505.56

**Write a QUERY TO FIND THE MONTHLY RETURN OF INVESTMENT (ROI) PER VOUCHER TYPE**

## Query

```
SELECT
    format_timestamp('%Y-%m', timestamp_event) month,
    b.voucher_type,
    round(sum(a.paid_amount)/sum(a.voucher_amount)) ROI
FROM `revoufsda-ecommerceeu.EcommerceU_B3.Transaction` a
    join `revoufsda-ecommerceeu.EcommerceU_B3.Voucher_Type` b
        on a.voucher_id = b.voucher_id
WHERE a.event_name = 'complete' and a.voucher = true
group by 1,2
order by 1;
```

## Result

Row	month	voucher_type	ROI
1	2019-01	cashback	26.0
2	2019-01	direct discount	30.0
3	2019-01	delivery fee discount	37.0
4	2019-02	cashback	31.0
5	2019-02	direct discount	28.0
6	2019-02	delivery fee discount	29.0
7	2019-03	cashback	27.0
8	2019-03	direct discount	28.0
9	2019-03	delivery fee discount	31.0
10	2019-04	cashback	34.0
11	2019-04	direct discount	33.0
12	2019-04	delivery fee discount	25.0
13	2019-05	cashback	34.0
14	2019-05	direct discount	30.0
15	2019-05	delivery fee discount	30.0

## WRITE A QUERY TO FIND VOUCHER UTILIZATION RATE FOR EACH MONTH

### Query

```
WITH voucher AS (
SELECT
    format_timestamp('%Y-%m', timestamp_event) month,
    count (distinct transaction_id) transaction,
    count (case when voucher=true then transaction_id else null end) transaction_voucher
FROM `revoufsda-ecommerceu.EcommerceU_B3.Transaction`
WHERE event_name = 'complete'
group by 1
order by 1
)

SELECT
month,
round(transaction_voucher/transaction,2) Voucher_utilization_rate
FROM voucher
ORDER BY 1;
```

### Result

Row	month	Voucher_utilization_rate
1	2019-01	0.48
2	2019-02	0.47
3	2019-03	0.58
4	2019-04	0.46
5	2019-05	0.5
6	2019-06	0.45
7	2019-07	0.55
8	2019-08	0.51
9	2019-09	0.48
10	2019-10	0.52
11	2019-11	0.55
12	2019-12	0.53
13	2020-01	0.54
14	2020-02	0.55
15	2020-03	0.53

## WRITE QUERY TO GET MONTH TO DATE (MtD) OF TOTAL REVENUE FOR EACH PROVINCE FOR TH PAST 3 MONTHS (CURRENT DATE 17 SEPTEMBER 2021), BREAKDOWN BY DATE (show: date; province, revenue)

### Query

```
with mtd as (
SELECT
    u.province AS province,
    DATE(timestamp_event) AS date,
    SUM(paid_amount) AS total_revenue
FROM
    `revoufsda-ecommerceu.EcommerceU_B3.Transaction` AS t
JOIN
    `revoufsda-ecommerceu.EcommerceU_B3.User` AS u
ON
    t.user_id = u.user_id
WHERE
    DATE(timestamp_event) between DATE_SUB('2021-06-01', INTERVAL 3 MONTH) and '2021-09-01'
    and extract(day from timestamp_event) between 1 and 17 and event_name = 'complete'
GROUP BY
    date;
```

```

1,2
ORDER BY
1,2)
select province, date,
sum(total_revenue) over (partition by province,date_trunc(date,month) order by date) as total
from mtd
order by 1,2;

```

## Result

Row	province	date	total
1	Banten	2021-03-01	2835000
2	Banten	2021-03-02	4055000
3	Banten	2021-03-03	5035000
4	Banten	2021-03-04	6785000
5	Banten	2021-03-05	7360000
6	Banten	2021-03-08	10645000
7	Banten	2021-03-09	11445000
8	Banten	2021-03-10	11475000
9	Banten	2021-03-12	13855000
10	Banten	2021-03-13	14560000
11	Banten	2021-03-14	16125000
12	Banten	2021-03-15	18375000
13	Banten	2021-03-17	18950000
14	Banten	2021-04-01	3550000
15	Banten	2021-04-02	3750000

Row	province	date	total
101	DKI Jakarta	2021-03-09	25465000
102	DKI Jakarta	2021-03-10	30615000
103	DKI Jakarta	2021-03-11	35830000
104	DKI Jakarta	2021-03-12	40815000
105	DKI Jakarta	2021-03-13	44350000
106	DKI Jakarta	2021-03-14	51000000
107	DKI Jakarta	2021-03-15	57645000
108	DKI Jakarta	2021-03-16	60390000
109	DKI Jakarta	2021-03-17	64470000
110	DKI Jakarta	2021-04-01	2435000
111	DKI Jakarta	2021-04-02	9520000
112	DKI Jakarta	2021-04-03	16040000
113	DKI Jakarta	2021-04-04	23640000
114	DKI Jakarta	2021-04-05	28815000
115	DKI Jakarta	2021-04-06	33510000

WRITE QUERY TO DO COHORT ANALYSIS BY REGISTERED MONTH

## Query

```

with cohort_items as (
    select
        user_id,
        MIN(date(date_trunc(register_date,MONTH))) as cohort_month,
        from `revoufsda-ecommerceeu.EcommerceU_B3.User`
        GROUP BY 1
),
user_activities as (
    select
        act.user_id as user_id,
        DATE_DIFF(
            date(date_trunc(timestamp_event,MONTH)),
            cohort.cohort_month,
            MONTH
        ) as month_number,
        from `revoufsda-ecommerceeu.EcommerceU_B3.Transaction` act
        left join cohort_items cohort ON act.user_id = cohort.user_id
        where extract(year from cohort.cohort_month) in (2019,2020,2021)
        group by 1, 2
),
cohort_size as (
    SELECT cohort_month,
        count(1) as num_users
        FROM cohort_items
        GROUP BY 1
        ORDER BY 1
),
retention_table as (
    select
        C.cohort_month,
        A.month_number,
        count(1) as num_users
        from user_activities A
        left join cohort_items C ON A.user_id = C.user_id
        group by 1, 2
)
-- our final value: (cohort_month, size, month_number, percentage)
select
    B.cohort_month,
    S.num_users as cohort_size,
    B.month_number,
    B.num_users as total_users,
    round((cast(B.num_users as decimal)/ S.num_users)*100,2)|| '%' as percentage
from retention_table B
left join cohort_size S ON B.cohort_month = S.cohort_month
where B.cohort_month IS NOT NULL
order by 1, 3

```

## Result

Row	cohort_month	cohort_size	month_number	total_users	percentage
1	2019-01-01	244	0	224	91.8%
2	2019-01-01	244	1	51	20.9%
3	2019-01-01	244	2	30	12.3%
4	2019-01-01	244	3	46	18.85%
5	2019-01-01	244	4	56	22.95%
6	2019-01-01	244	5	50	20.49%
7	2019-01-01	244	6	42	17.21%
8	2019-01-01	244	7	36	14.75%
9	2019-01-01	244	8	49	20.08%
10	2019-01-01	244	9	59	24.18%
11	2019-01-01	244	10	54	22.13%
12	2019-01-01	244	11	51	20.9%
13	2019-01-01	244	12	55	22.54%
14	2019-01-01	244	13	56	22.95%
15	2019-01-01	244	14	40	16.39%