

```
### MASTER_PROJECT--1 (Globox-A/B Testing)
```

```
/* How many users in the control group were in  
Canada? This question is required.*/
```

```
SELECT count(*)  
FROM users as u  
JOIN groups as g  
ON  
u.id = g.uid  
WHERE country = 'CAN' AND "group" = 'A';
```

```
/* What was the conversion  
rate of all users? This question is required.*/
```

```
SELECT  
count(distinct a.uid)*100 /  
count(distinct u.id) :: NUMERIC AS conversion_rate  
FROM activity a  
RIGHT JOIN users as u  
ON  
a.uid = u.id
```

```
/*As of February 1st, 2023, how many users were in the A/B test?*/
```

```
SELECT  
COUNT(*) AS total_users  
FROM groups  
WHERE join_dt <= '2023-02-01'
```

```
/* What is the  
average amount spent per user for the control and treatment groups?*/
```

```
SELECT g.group,  
SUM(COALESCE(a.spent,0))/COUNT(DISTINCT(u.id)) AS avg_spent_per_user  
FROM activity AS a  
RIGHT  
JOIN groups AS g  
ON a.uid = g.uid  
INNER JOIN users as u  
ON g.uid = u.id  
WHERE g.group IN  
( 'A', 'B' )  
GROUP BY g.group;  
----- OR -----  
SELECT g.group, ROUND(CAST(SUM(COALESCE(spent,  
0))/COUNT(DISTINCT g.uid) AS numeric),3) as average  
FROM groups as g  
LEFT JOIN activity AS  
a  
USING (uid)  
GROUP BY g.group
```

```
/*What is the 95% confidence interval for the average  
amount spent per user in the control? Use the t distribution.*/
```

```
WITH cte AS  
( SELECT uid,  
"group", SUM(spent) AS total_spent  
FROM groups  
LEFT JOIN  
activity  
USING(uid)  
GROUP BY uid, "group" ),  
cte_2 AS  
(SELECT uid,  
"group", (COALESCE(total_spent, 0)) total_spent  
FROM cte)  
SELECT "group",  
AVG(total_spent) AS mean_spending,  
STDDEV(total_spent) AS standard_deviation,  
COUNT(distinct uid) AS sample_size,  
AVG(total_spent) - 1.96 * STDDEV(total_spent) /  
SQRT(COUNT(uid)) AS lower_bound,
```

```

        AVG(total_spent) + 1.96 * STDDEV(total_spent) /
SQRT(COUNT(uid)) AS upper_bound
FROM cte_2
GROUP BY "group";

```

/*What is the 95% confidence interval for the average amount spent per user in the treatment? Use the t distribution.*/

```

WITH cte AS
( SELECT uid, "group", SUM(spent) AS
total_spent
FROM groups
LEFT JOIN activity
USING(uid)
GROUP BY uid, "group"
),
cte_2 AS
(SELECT uid, "group", (COALESCE(total_spent, 0)) total_spent
FROM
cte)
SELECT "group", AVG(total_spent) AS mean_spending,
STDDEV(total_spent)
AS standard_deviation, COUNT(distinct uid) AS sample_size,
AVG(total_spent) - 1.96 *
STDDEV(total_spent) / SQRT(COUNT(uid)) AS lower_bound,
AVG(total_spent) + 1.96 *
STDDEV(total_spent) / SQRT(COUNT(uid)) AS upper_bound
FROM cte_2
GROUP BY
"group";

```

/* Conduct a hypothesis test to see whether there is a difference in the average amount spent per user between the two groups. What are the resulting p-value and conclusion? Use the t distribution and a 5% significance level. Assume unequal variance.*/

```

WITH cte AS
( SELECT uid, "group", SUM(spent) AS total_spent
FROM groups

LEFT JOIN activity
USING (uid)
GROUP BY uid,"group"),
cte_2 AS

(SELECT uid, "group", COALESCE(total_spent, 0) AS total_spent
FROM cte)
SELECT
* FROM cte_2;
----- OR-----
WITH cte AS
(SELECT uid, "group",SUM(spent)AS
total_spent,
CASE WHEN SUM(spent) is null THEN 0
WHEN SUM(spent)=0 THEN 0
WHEN
SUM(spent)= 0 THEN 1 END AS conversion
FROM groups
LEFT JOIN activity
USING (uid)
GROUP BY
uid,"group"),
cte_2 AS
(SELECT uid,"group",(COALESCE(total_spent,0))
total_spent,conversion
FROM cte)
SELECT * from cte_2

```

We used this code to download the required columns as csv file to further process the query in google sheets.

/* What is the user conversion rate for the control and treatment groups?*/

```
SELECT
g.group,ROUND(COUNT(DISTINCT a.uid)/COUNT(DISTINCT u.id):: NUMERIC *100,2) AS
conversion_rate
FROM users AS u
JOIN groups AS g
ON u.id = g.uid
LEFT JOIN activity AS
a
ON g.uid = a.uid
WHERE g.group IN ('B','A')
GROUP BY 1
```

```
/**/
```