GreatPlants XP: replace “Let’s Go” button with a more compact sign-up form.

*Last updated: **Apr 03, 2023**, Author: Xinyu Li*

Section 01 Summary *(write at the end of the experiment)*

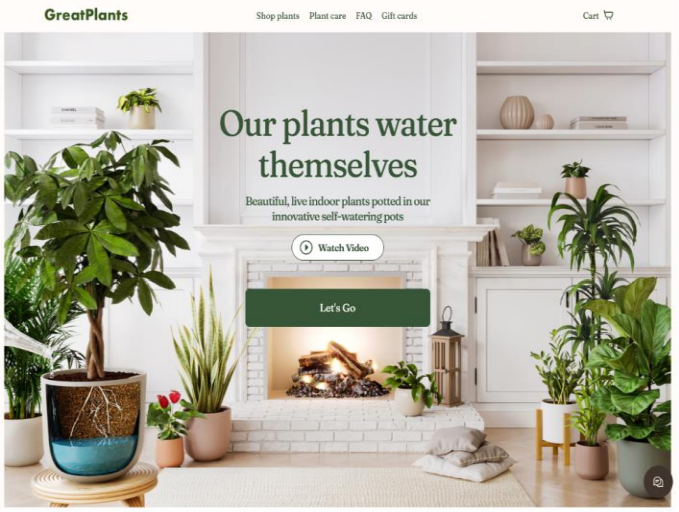
*We wrote this section on **April 03, 2023** one day after the experiment was completed.*

- A new version of the landing page with a more compact signup flow was developed by the marketing team.
- We ran an AB test of 2 weeks, between 2023-03-20 and 2023-04-02, on **1200** leads.
- The results show a positive impact with statistical significance on:
    - primary metric: percentage of leads becoming registered users, with an increase of 0.123909.
- There is no negative impact with statistical significance on secondary and guardrail metrics.
- Given that there are no significant negative effects on secondary and guardrail metrics, and the new feature has a significant positive impact on the primary metric: percentage of leads becoming registered users. We recommend rolling-out the feature.

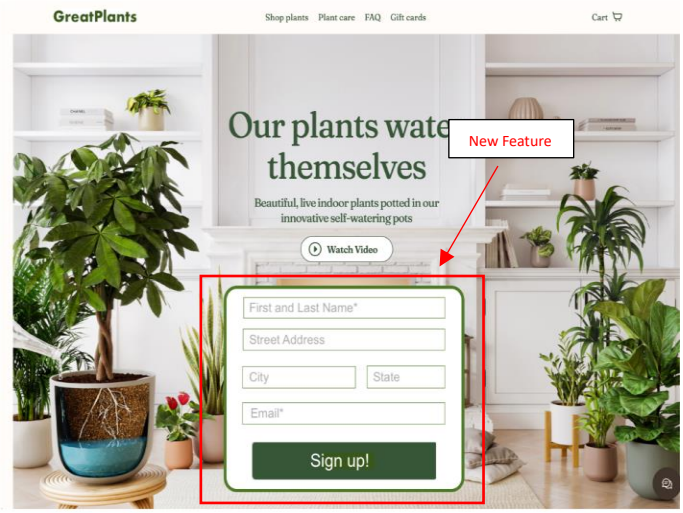
Section 02 Description of the feature and experiment

*Last updated: **Mar 19, 2023**, Author: Xinyu Li, before running the experiment*

Our marketing team has decided to revamp the landing page with a more compact signup flow. Rather than using the "Let's Go" button to redirect leads to a complex registration process, the new version will allow new customers to input their information directly on the landing page to sign up. To optimize the user experience, we plan to simplify the form and use clear, easy-to-understand language.



Control

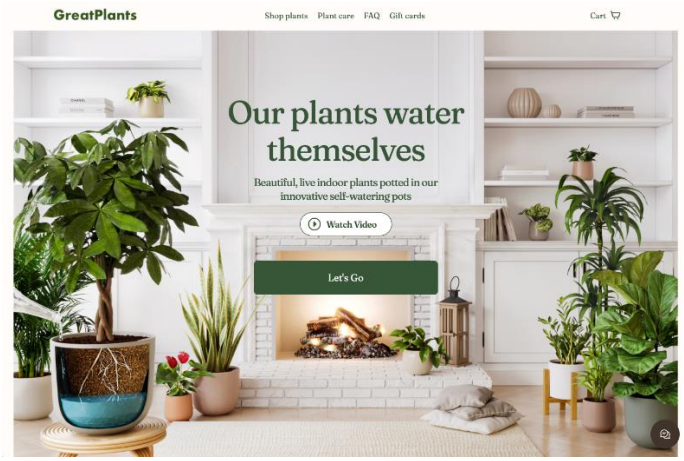


Treatment

The main idea is that we want to build a strong and long-lasting relationship with potential customers. To achieve this, we are focusing on converting new leads into registered clients (increase sign-up rate on the landing page). Hopefully, this new feature will encourage more leads to sign up and improve key performance metrics (defined in Section 03.d Success Metric).

The experiment is a classic AB experiment:

- For a period of two weeks, **1200** leads (defined in Section 03.e Power Analysis) will be split 50/50 into 2 groups: Control and Treatment.
- The Control group will see the old version of landing page with “Let’s Go” button redirecting them to a complex sign-up process.



- The Treatment group will see the new version of landing page with a compact sign-up form instead of a “Let’s Go” bottom.

After two weeks we will analyze the data and we will determine that either:

- The new feature has a significant impact on key performance metrics and should be rolled out.
- Or the new feature does not show a significant improvement. In this case, we will not roll out the new feature.

Section 03 Experiment Planning

*Last updated: **Mar 19, 2023**, Author: Xinyu Li, before running the experiment*

In this section we will cover platforms, unit of randomization, groups, success metrics, power analysis, data logging, and any other details that are needed to run the experiment.

Section 03.a platform(s)

A new feature that replaces a “Let’s Go” button with a more compact sign-up process, will appear (for leads in Treatment group) on the landing page. No other platforms (web pages) will be affected.

Section 03.b Unit of randomization

The experiment will be administered at the lead level. It is reasonable to assume that leads behave independently from each other on our platform, and there will be no contamination between the Control group and the Treatment group.

The alternative option would be to administer the experiment at the session level (sessions being randomized in Control and Treatment). However, this would not be a good idea: leads can have multiple sessions (therefore sessions would not be independent) and the same lead might see both versions of our website, creating a confusing experience.

Section 03.c Control/Treatment groups

The exact number of leads entering the experiment (either in Control or Treatment) will be around 1200, which is calculated and determined in the Power Analysis section (Section 03.e Power Analysis).

50% of these leads will be randomly assigned to the Control group (they will see a "Let's Go" button instead of a more compact sign-up form in the landing page), and the other 50% will be assigned to the Treatment group (they will see a more compact sign-up form instead of a "Let's Go" button in the landing page).

Notice that during the experiment, new leads (that have never been on the website before) will start their session and at the time they will be randomized in the Control and Treatment groups.

Section 03.d Success Metrics

We have defined the following success metrics that we will compute at the lead level, over the period of 2 weeks. To get an idea of the baselines, we considered

- the two weeks between 2023-03-06 and 2023-03-19
- the 1,200 new leads (that have never been on the website before) started their first session on our website in those two weeks.

We computed the following metrics based on all leads during two weeks between 2023-03-06 and 2023-03-19: (1) "percentage of new leads becoming registered users"; (2) "average price paid during the two-week experiment before tax per lead"; (3) "average number of orders placed directly from the landing page per new lead"; (4) "price before tax per order".

We compute the following metrics based on a portion of leads who signed up during two weeks between 2023-03-06 and 2023-03-19: "average time took from first time landing the page to sign-up in hours per sign-up lead".

Metric	Type	Priority of metric	Default value	Desired effect
percentage of leads becoming registered users	sample mean	primary	0.454167	go up

average time took from first time landing the page to sign-up in hours per sign-up lead	sample mean	secondary	56.337615	go down
average price paid during the two-week experiment before tax per lead	sample mean	secondary	98.220092	go up
average number of orders placed directly from the landing page per lead	sample mean	guardrail	0.590833	no effect
price before tax per order	Ratio metric (average price paid before tax per lead / average number of orders per lead)	guardrail	119.658995	no effect

(See Query 01 – Query 05 in Appendix for detailed SQL query calculations)

Section 03.e Power Analysis

We want to run the experiment for 2 weeks (the week between 2023-03-20 and 2023-04-02). We have hundreds of thousands of leads on our platform but won't need to run the experiment on all of them. A subset will suffice.

We will base power analysis on the primary metric, looking at data from the two weeks right before the experiment, these are the two weeks between 2023-03-06 and 2023-03-19.

As seen in the table above, on average, in the two weeks between 2023-03-06 and 2023-03-19, percentage of leads becoming registered users is about 0.4542. The standard deviation (sigma hat), estimated from the data is 0.4982 (given by Query 1 in the appendix).

Let's fix alpha (probability of Type I error) to be 5%. Actually, given that we are going to run a test on 5 metrics, we will use the Bonferroni correction to be conservative. Therefore, our alpha value is 0.05/5 = 0.01.

Let's fix beta (probability of Type II error) to be 20%.

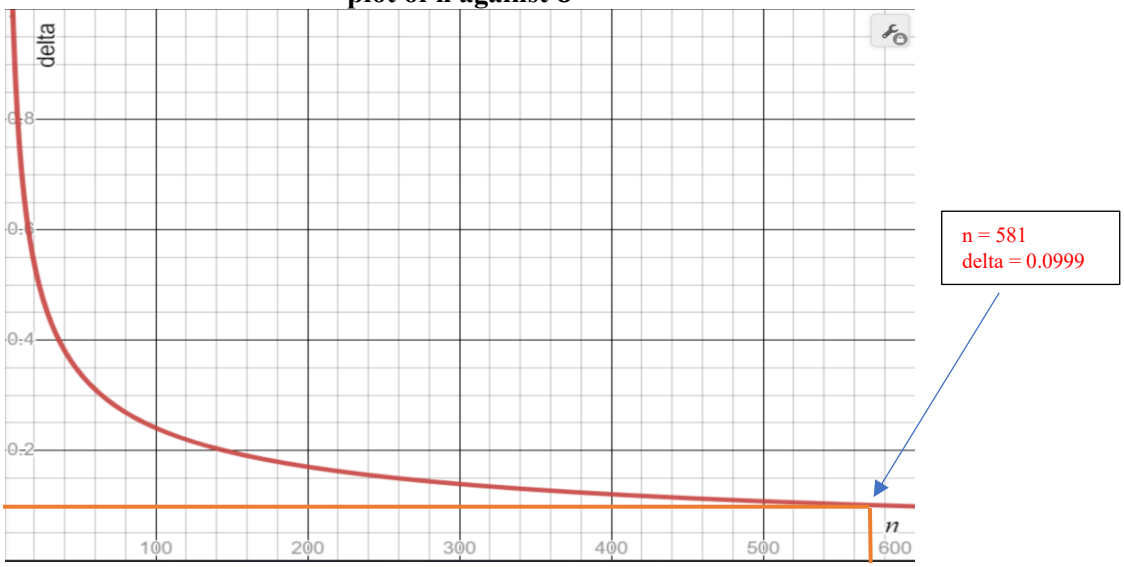
Based on the given information, we can find the relationship between δ (difference between treatment group and control group) and n (sample size) as follow:

relationship between δ and n

$$n = 2\times(\hat{\sigma}\times\frac{Z_{\alpha/2}+Z_{\beta}}{\delta})^2 = 2\times(0.4981\times\frac{Z_{0.01/2}+Z_{0.2}}{\delta})^2 = 2\times(0.4981\times\frac{2.5758+0.8416}{\delta})^2 = \frac{5.795}{\delta^2}$$

The associated graph of n (on the x-axis) against δ (on the y-axis) is also plotted below:

plot of n against δ



After discussing with the marketing team, we expect the new feature to have an impact on the percentage of leads becoming registered users of at least 22% of the baseline, which is equivalent to 22% * 0.4542 = 0.0999. The corresponding desired sample size is calculated as followings:

$$n = \frac{5.795}{\delta^2} = \frac{5.795}{0.0999^2} \approx 581$$

Based on our calculation there is about 581 leads required in each group. Rounding up the numbers (to be conservative, since there is always some spillage), we believe we can run an experiment on 1200 (2*600 = 1200) leads (50% in treatment, 50% in control). If the difference in the primary metric will be at least 0.0999 there is a high probability (80%) that our experiment will detect it.

Section 03.f Data Logging

During the experiment, new leads (that have never been on the website before) will start their first session and at that time our Experimentation platform will automatically store the allocation of leads in Control and Treatment in the table **experiment_groups**, which contains the following columns:

- experiment_name: the unique identifier of an experiment. we will call our experiment "landing_2023".
- unit_id: the identifier of the unit of randomization that was used in the experiment. in our case this will correspond to "lead_id".
- grouped_at: the timestamp of when a lead was assigned to a group in an experiment.
- group: the name of the group that a lead was assigned to.

We won't have to worry about logging additional data, since standard business tables will log data for the behavior of the leads. In particular:

- **dim_lead** contains information about each lead that ever visited a landing page. There is one row per lead that ever started a session on a landing page. Note that a lead can have multiple sessions. The associated columns are shown as follows:
    - lead_id: the lead identifier.
    - lead_source: can be one of direct / search / ads / social.
    - first_session_at: timestamp of the first session of a lead.
    - signup_at: timestamp of the signup completion. If NULL, the lead never signed up.

    - lead_info: encrypted information provided by the lead during signup (name, address, email, etc). If NULL, the lead never signed up.
- **fact_landing** contains information about each website session that starts on a landing page. There is one row per session. Note that it does not contain sessions that start somewhere else other than landing page, and one lead can have multiple sessions, and it's possible that customers do not place order directly from the landing page, which may cause order_id to be null. The associated columns are shown as follows:
    - lead_id: the lead identifier.
    - session_id: the session identifier.
    - session_at: timestamp of the session.
    - signup_at: if a signup was completed during the session, the timestamp of the signup is recorded here.
    - order_id: if an order was placed during the session, the order_id is recorded here.
- **fact_order** contains information about every order, regardless of the flow that generated it. There is one row per order, which could be generated from a session that started on a landing page or from a different flow. The associated columns are shown as follows:
    - lead_id: the lead or client identifier (even after a lead signs up, we still identify them with their lead_id).
    - order_id: the order identifier.
    - order_at: timestamp of the order.
    - tot_price: total price of the order, before taxes.
    - taxes: sales tax applied to the order.
    - tot_cost: total cost paid by the customer, including taxes.

Section 04 Rollout

*We wrote this section on **Mar 20, 2023**, at the beginning of the Experiment.*

- On 2023-03-19 everything was in place to start the experiment. We used this day to start with a slow roll-out (1% and 10%) and monitored the systems. Everything was ok. In particular the allocation of leads in Control and Treatment was approximately 50/50.
- On 2023-03-20 we officially rolled-out to our 1200 leads (100%), which started to enter in the experiment. The platform automatically logs the corresponding data.

Here is a screenshot of our internal experimentation platform and the corresponding parameters that we set in the system, in partnership with the Engineering team:

replace "Let's Go" button with a more compact sign-up form	Name of the experiment
lead_id	Unit of randomization
Web UX	Platforms
2023-03-20 00:00:00 UTC	XP start timestamp
2023-04-02 23:59:59 UTC	XP end timestamp
two	Number of variants
default	First variant
change "Let's Go" button to a compact sign-up flow	Second variant
1%, 10%, 100% - manual	XP rollout mode

Section 05 Analysis of results

*We wrote this section on **April 03, 2023** one day after the experiment was completed.*

The experiment ran between 2023-03-20 and 2023-04-02.

We saw:

- 598 active leads in the Control group
- 603 active leads in the Treatment group

We don't see 600 leads in each group because of some minor spillage (data that was not logged correctly). The 50/50 randomization has worked well since we see almost exactly 50% in Control and 50% in Treatment.

We now have all the data to run our analysis. We will consider all the metrics in the "Success metric" section above, and we will compare the values obtained in Control and Treatment with statistical tests. Given that we will perform 5 tests, we will use the Bonferroni correction to declare a result statistically significant. That is, instead of setting a threshold of 0.05 for our p-values, the threshold is 0.05/5 = 0.01.

Query 6 to Query 10 in the appendix return the data that we used in "P-Value Calculation" tab in "SQL Final Spreadsheet" to do our calculations, summarized in the table below:

Metric	**Control group**	**Treatment group**	**Difference (T-C)**	**p-value**
percentage of leads becoming registered users	0.456522	0.580431	0.123909	0.000015
average time took from first time landing the page to sign-up in hours per sign-up lead	52.318681	37.660000	-14.658681	0.016469
average price paid during the two-week experiment before tax per lead	102.247425	110.933930	8.686505	0.183227
average number of orders placed directly from the landing page per lead	0.652174	0.590381	-0.061793	0.124991
price before tax per order	120.3621302	120.5282171	0.166087	0.787114

Section 06 Conclusion

*We wrote this section on **April 03, 2023** one day after the experiment was completed.*

The A/B test results indicate that the new feature of replacing the "Let's Go" button with a more compact sign-up form on the landing page has had a statistically significant positive impact on the primary metric: "the percentage of leads becoming registered users", with an increase of 0.123909 (p-value=0.000015). What's more the increased amount exceeds the threshold set by the marketing team: 0.0999.

Although the new feature has a positive impact on other metrics such as: (1) average time took from first time landing the page to sign-up in hours per sign-up lead, with a decrease of 14.658681 (2) average price paid during the two-week experiment before tax per lead, with an increase of 8.686505 (3) price before

tax per order, with an increase of 0.166087, these effects are not statistically significant, and they will be ignored from our analysis.

There is a negative effect on “average number of orders placed directly from the landing page per lead”. However, this is not statistically significant either. Thus, it will be ignored from our analysis.

Overall, the primary metric of percentage of leads becoming registered users has shown a significant positive impact and exceeded the market team’s expectation, and there are no other significant negative effects on the secondary and guardrail metrics. Therefore, **it is recommended to roll-out the new feature to all leads**, as it has shown to have a significant positive impact on the user experience and business outcomes.

Section 07 Suggested Future Work

*We wrote this section on **April 03, 2023** one day after the experiment was completed.*

Suggested future work 01:

In this A/B test experiment, only the primary metric has shown a statistically significant difference between the control and variant groups, but the secondary and guardrail metrics have not, it may be necessary to conduct further A/B testing experiments with a larger sample size to better understand the effect of the new feature on these metrics. The statistically insignificant results for the secondary and guardrail metrics could be due to insufficient power or random variation, and increasing the sample size could help to reduce the impact of these factors and improve the statistical power of the experiment. With a larger sample size, it may be possible to identify any significant differences in the secondary and guardrail metrics and determine the overall impact of the new feature on the user experience.

Suggested future work 02:

Metric “average time took from first time landing the page to sign-up in hours per sign-up lead” in this A/B test only uses a subset of the leads (273 registered leads in control group and 350 registered leads in treatment group), specifically those who have signed up, it may be necessary to redo the experiment with an increased sample size to address this issue. Using only a portion of the leads may reduce the statistical power of the experiment, and increasing the sample size could help to ensure that there are enough sign-up leads to meet the minimum amount required in the power analysis. This will enable a more accurate analysis of the data and provide a better understanding of the impact of the new feature on “average time took from first time landing the page to sign-up in hours per sign-up lead”.

Suggested future work 03:

It is possible that the new feature of replacing the "Let's Go" button with a more compact sign-up form may have a novelty effect on certain metrics, such as the “average price paid during the two-week experiment before tax per lead”. To account for this potential effect and to ensure the accuracy and validity of the results, it is recommended to conduct another A/B test that is similar to the previous experiment but with a longer duration.

In this new A/B test, it is suggested to ignore orders placed within the first 5 days after a new lead first launches the website, in order to minimize the impact of the novelty effect. By doing so, the results of the experiment will be more reliable and accurate, and any observed effects will be more likely to reflect the true impact of the new feature on the user behavior and business outcomes.

Overall, conducting another A/B test with a longer duration and adjusted criteria is a good approach to confirm the findings and evaluate the impact of the new feature more accurately, especially on metrics that may be influenced by the novelty effect.

Appendix

Query 01: Percentage of new leads that become registered users per lead (before experiment)

Codes:

```
WITH dim_lead_modified AS(
  SELECT *,CASE WHEN (signup_at >= CAST("2023-03-
06 00:00:00 UTC" as TIMESTAMP) AND signup_at < CAST("2023-03-
19 23:59:00 UTC" as TIMESTAMP)) THEN 1 ELSE 0 END AS if_signed_up
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
)

SELECT COUNT(*) AS sample_size, STDDEV(if_signed_up) AS standard_deviation, AVG(if_sig
ned_up) AS mean
FROM dim_lead_modified
```

Results:

Row	sample_size	standard_deviati	mean
1	1200	0.49810245...	0.45416666...

Query 02: average time took from first time landing the page to sign-up in hours per sign-up lead (before experiment)

Codes:

```
WITH dim_lead_modified AS(
  SELECT *
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
  AND signup_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND signup_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
)

SELECT AVG(DATE_DIFF(signup_at,first_session_at,HOUR)) AS mean
FROM dim_lead_modified d
WHERE d.signup_at IS NOT NULL
```

Results:

Row	mean
1	56.3376146...

Query 03: average price paid during the two-week experiment before tax per lead (before experiment)

Codes:

```
WITH customers_within_time_period AS(
  SELECT DISTINCT lead_id, first_session_at
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
),
all_orders_1 AS(
  SELECT *
  FROM fabriziopublic.greatplants.fact_order
  WHERE 1=1
  AND order_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND order_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
),
all_orders_2 AS(
  SELECT t2.lead_id, t1.order_id, COALESCE(t1.tot_price,0) AS tot_price
  FROM all_orders_1 t1 RIGHT JOIN customers_within_time_period t2 ON t1.lead_id = t2.l
ead_id
),
all_orders_3 AS(
  SELECT lead_id, sum(tot_price) as total_order_value
  FROM all_orders_2
  GROUP BY 1
)
SELECT AVG(total_order_value) AS mean
FROM all_orders_3
```

Results:

Row	mean
1	98.2200916…

Query 04: average number of orders placed directly from the landing page per lead (before experiment)

Codes:

```
WITH first_time_cust AS(
  SELECT DISTINCT lead_id, first_session_at
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
),
landing_page_order AS(
  SELECT lead_id, SUM(IF(order_id IS NULL,0,1)) AS total_num_orders
  FROM fabriziopublic.greatplants.fact_landing
  WHERE session_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND session_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
  GROUP BY 1
),
tot_landing_page_oder_per_user AS(
  SELECT t1.lead_id, COALESCE(total_num_orders,0) AS tot_num_orders
  FROM first_time_cust t1 LEFT JOIN landing_page_order t2 ON t1.lead_id = t2.lead_id
)

SELECT AVG(tot_num_orders) AS mean
FROM tot_landing_page_oder_per_user
```

Results:

Row	mean
1	0.59083333...

Query 05: price before tax per order (before experiment)

Codes:

```
WITH customers_within_time_period AS(
  SELECT DISTINCT lead_id, first_session_at
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
), all_orders_1 AS(
  SELECT *
  FROM fabriziopublic.greatplants.fact_order
  WHERE 1=1
  AND order_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND order_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
), all_orders_2 AS(
  SELECT t2.lead_id, t1.order_id, COALESCE(t1.tot_price,0) AS tot_price
  FROM all_orders_1 t1 RIGHT JOIN customers_within_time_period t2 ON t1.lead_id = t2.l
ead_id
), all_orders_3 AS(
  SELECT lead_id, sum(tot_price) as total_order_value
  FROM all_orders_2
  GROUP BY 1
), all_orders_4 AS(
  SELECT *
  FROM fabriziopublic.greatplants.fact_order
  WHERE 1=1
  AND order_at >= CAST("2023-03-06 00:00:00 UTC" as TIMESTAMP)
  AND order_at < CAST("2023-03-19 23:59:00 UTC" as TIMESTAMP)
), all_orders_5 AS(
  SELECT t2.lead_id, IF(t1.order_id IS NULL, 0 , 1) AS num_order
  FROM all_orders_4 t1 RIGHT JOIN customers_within_time_period t2 ON t1.lead_id = t2.l
ead_id
), all_orders_6 AS(
  SELECT lead_id, SUM(num_order) AS total_num_order
  FROM all_orders_5
  GROUP BY 1
), avg_num_order AS(
  SELECT AVG(total_num_order) AS avg_order
  FROM all_orders_6
), avg_order_value AS(
  SELECT AVG(total_order_value) AS avg_value
  FROM all_orders_3
)
SELECT (SELECT avg_value FROM avg_order_value)/(SELECT avg_order FROM avg_num_order) A
S price_per_order
```

Results:

Row	price_per_order
1	119.658994...

Query 06: percentage of new leads becoming registered users (after experiment)

Codes:

```
WITH dim_lead_modified AS(
  SELECT *,CASE WHEN (signup_at >= CAST("2023-03-
20 00:00:00 UTC" as TIMESTAMP) AND signup_at < CAST("2023-04-
02 23:59:00 UTC" as TIMESTAMP)) THEN 1 ELSE 0 END AS if_signed_up
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
)
SELECT t2.group, COUNT(*) AS sample_size, STDDEV(if_signed_up) AS standard_deviation,
AVG(if_signed_up) AS mean
FROM dim_lead_modified t1 RIGHT JOIN fabriziopublic.greatplants.experiment_groups t2 O
N t1.lead_id = t2.unit_id
where t2.experiment_name = "landing_2023"
GROUP BY 1
```

Results:

Row	group	sample_size	standard_deviati	mean
1	Control	598	0.49852305…	0.45652173…
2	Treatment	603	0.49389812…	0.58043117…

Query 07: average time took from first time landing the page to sign-up in hours per sign-up lead (after experiment)

Codes:

```
WITH dim_lead_modified AS(
  SELECT *,CASE WHEN signup_at IS NULL THEN 0 ELSE 1 END AS if_signed_up
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
  AND signup_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND signup_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
)

SELECT t2.group, COUNT(*) AS sample_size, AVG(DATE_DIFF(signup_at,first_session_at,HOU
R)) AS mean, STDDEV(DATE_DIFF(signup_at,first_session_at,HOUR)) AS standard_deviation
FROM dim_lead_modified t1 JOIN fabriziopublic.greatplants.experiment_groups t2 ON t1.l
ead_id = t2.unit_id
WHERE t1.signup_at IS NOT NULL
AND t2.experiment_name = "landing_2023"
GROUP BY t2.group
```

Results:

Row	group	sample_size	mean	standard_deviati
1	Control	273	52.3186813…	79.7853571…
2	Treatment	350	37.6600000…	70.0974462…

Query 08: average price paid during the two-week experiment before tax per lead (after experiment)

Codes:

```
WITH customers_within_time_period AS(
  SELECT DISTINCT lead_id, first_session_at
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
),
all_orders_1 AS(
  SELECT *
  FROM fabriziopublic.greatplants.fact_order
  WHERE 1=1
  AND order_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND order_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
),
all_orders_2 AS(
  SELECT t2.lead_id, t1.order_id, COALESCE(t1.tot_price,0) AS tot_price
  FROM all_orders_1 t1 RIGHT JOIN customers_within_time_period t2 ON t1.lead_id = t2.l
ead_id
),
all_orders_3 AS(
  SELECT *
  FROM all_orders_2 t1 RIGHT JOIN fabriziopublic.greatplants.experiment_groups t2 ON t
1.lead_id = t2.unit_id
  WHERE t2.experiment_name = "landing_2023"
),
all_orders_4 AS(
  SELECT unit_id, a.group, sum(tot_price) as total_order_value
  FROM all_orders_3 a
  GROUP BY 1,2
)
SELECT a.group, COUNT(*) AS sample_size, AVG(total_order_value) AS mean, STDDEV(total_
order_value) AS standard_deviation
FROM all_orders_4 a
GROUP BY 1
```

Results:

Row	group	sample_size	mean	standard_deviati
1	Control	598	102.247424...	109.146458...
2	Treatment	603	110.933930...	116.940910...

Query 09: average number of orders placed directly from the landing page per lead (after experiment)

Codes:

```
WITH first_time_cust AS(
  SELECT DISTINCT lead_id, first_session_at
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
),
landing_page_order AS(
  SELECT lead_id, SUM(IF(order_id IS NULL,0,1)) AS total_num_orders
  FROM fabriziopublic.greatplants.fact_landing
  WHERE session_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND session_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
  GROUP BY 1
),
tot_landing_page_order_per_user AS(
  SELECT t1.lead_id, COALESCE(total_num_orders,0) AS tot_num_orders
  FROM first_time_cust t1 LEFT JOIN landing_page_order t2 ON t1.lead_id = t2.lead_id
)

SELECT t2.GROUP, AVG(t1.tot_num_orders) AS mean, STDDEV(t1.tot_num_orders) AS standard
_deviation
FROM tot_landing_page_order_per_user t1 RIGHT JOIN fabriziopublic.greatplants.experime
nt_groups t2 ON t1.lead_id = t2.unit_id
WHERE t2.experiment_name = "landing_2023"
GROUP BY 1
```

Results:

Row	GROUP	mean	standard_deviati
1	Control	0.65217391...	0.73629541...
2	Treatment	0.59038142...	0.65696398...

Query 10: price before tax per order (after experiment)

Codes:

```
WITH first_time_cust AS(
  SELECT DISTINCT lead_id, first_session_at
  FROM fabriziopublic.greatplants.dim_lead
  WHERE 1=1
  AND first_session_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND first_session_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
),
landing_page_order AS(
  SELECT lead_id, COUNT(*) AS total_num_orders
  FROM fabriziopublic.greatplants.fact_landing
  WHERE session_at >= CAST("2023-03-20 00:00:00 UTC" as TIMESTAMP)
  AND session_at < CAST("2023-04-02 23:59:00 UTC" as TIMESTAMP)
  GROUP BY 1
),
tot_landing_page_order_per_user AS(
  SELECT t1.lead_id, COALESCE(total_num_orders,0) AS tot_num_orders
  FROM first_time_cust t1 LEFT JOIN landing_page_order t2 ON t1.lead_id = t2.lead_id
)

SELECT t2.GROUP, AVG(t1.tot_num_orders) AS mean, STDDEV(t1.tot_num_orders) AS standard
_deviation
FROM tot_landing_page_order_per_user t1 RIGHT JOIN fabriziopublic.greatplants.experime
nt_groups t2 ON t1.lead_id = t2.unit_id
WHERE t2.experiment_name = "landing_2023"
GROUP BY 1
```

Results:

Row	group	mean_value	std_value	mean_num	std_num	covariance
1	Control	102.247424...	109.146458...	0.84949832...	0.90268058...	98.1908010...
2	Treatment	110.933930...	116.940910...	0.92039800...	0.96636035...	112.602024...