A/B Test conversion Rate Optimization (CRO)

Digital Analysis Project¶

Description: The product owner wants to improve the registration journey for a online game users. He has an idea that splitting the manual sign-up form from 1 step to 3 steps will improve the sign-up conversion rate. He is also interested in the impact of this change on purchasing behaviour (first purchase).

The data provided below was collected during the experiment using two variants of the sign-up form, Variant 1 (original): 25991400310 and Variant 2 (multistep): 25974540423. My task is to find out if the conversion rate has improved and to check whether there was an improvement in first purchase conversion and their value.

Data:
Table one - decisions

- experiment_id - unique experiment ID
    - set to NULL if the user is evaluated for the experiment and fails the audience condition
- variation_id - unique variation ID
  ○ set to NULL if the user is evaluated for the experiment and fails the audience condition
  ○ 25991400310 - original
  ○ 25974540423 - multistep
- visitor_id - hashed visitor id
- session_id - hashed session id
- date - creation date of data
- timestamp - decision timestamp in milliseconds (in UTC)
- user_agent - user-agent

Table two - conversions

- experiment_id - unique experiment ID
    - set to NULL if the user is evaluated for the experiment and fails the audience condition
- variation_id - unique variation ID
    - set to NULL if the user is evaluated for the experiment and fails the audience condition
    - 25991400310 - original
    - 25974540423 - multistep
- visitor_id - hashed visitor id
- session_id - hashed session id
- user_account_id - hashed user account id
- date - creation date of data
- timestamp - conversion timestamp in milliseconds (in UTC)
- event_type - event type (click, pageview, custom, or client_activation)
- event_name - Friendly event name
    - 25145940571_register_page_and_confirm
    - signUp - fires when user clicks “Play Now” button (before email confirmation)

Table three - account_details

- account_id - hashed user account id
    - (account_details.account_id=conversions.user_account_id)
- sign_up_timestamp - account creation timestamp
- first_purchase_timestamp - first purchase timestamp

- sign_up_method - registration method: manual (through our form) or using 3rd party account (e.g. google, facebook)
- first_purchase_value - total first purchase value

Objectives

To help the product owner makes a decision, we need to identify two elements:

1. To determine the conversion rate, we compare the number of visitors who signed up to the total number of visitors exposed to each variant of the experiment, in both the original and multistep forms.
2. To find the proportion of first-time purchasers compared to signed-up visitors in both variants. Additionally calculate the total first purchase value for both variants.

1. Determine the Sign-Up Conversion Rate

The sign-up conversion rate can be calculated as the number of unique users who completed the sign-up process (from conversions table) divided by the number of unique users who were exposed to each variation of the experiment (from decisions table).

Explanation:

- **decisions** and **conversions** tables are joined on `visitor_id` and `session_id`.
- The number of visitors exposed to each variation is counted.
- The number of visitors who completed the sign-up process (`event_name = 'signUp'`) is counted.
- The sign-up conversion rate is calculated as the ratio of sign-ups to the number of visitors.

SQL Query to Calculate Sign-Up Conversion Rate can be found here

```
-- Calculate the sign-up conversion rate for each variation

WITH distinct_visitors AS (

	SELECT
		d.variation_id,
		COUNT(DISTINCT d.visitor_id) AS distinct_visitor_count
	FROM
		decisions d
	GROUP BY
		d.variation_id
),

signups AS (
	SELECT
		c. variation_id,
		COUNT(DISTINCT c.visitor_id) AS signup_count
	FROM
		conversions c
	where
```

```
            c.event_name = 'signUp'
        group by
        c.variation_id
)

SELECT
    dv.variation_id,
    dv.distinct_visitor_count,
    ms.signup_count,
    (ms.signup_count * 1.0 / dv.distinct_visitor_count) AS conversion_rate
FROM
    distinct_visitors dv
JOIN
    signups ms ON dv.variation_id = ms.variation_id
;
```

Results:

variation_id	Total manual Visitors	sign_ups	conversion_rate
25974540423	7642	3934	0.514787
25991400310	7721	3940	0.510297

Significance calculations

To conclude whether the difference in conversion rate is statistically significant or not, we calculate z-score and p value for the result. Python was used as listed:

```
import scipy.stats as stats
import math

# Given data
n1 = 7721 # number of visitors for the original form
n2 = 7642  # number of visitors for the multi-step form
p1 = 0.510297 # conversion rate for the original form
p2 = 0.514787 # conversion rate for the multi-step form

# Calculate number of conversions
x1 = n1 * p1
x2 = n2 * p2

# Calculate pooled proportion
p_pooled = (x1 + x2) / (n1 + n2)

# Calculate standard error
SE = math.sqrt(p_pooled * (1 - p_pooled) * (1/n1 + 1/n2))

# Calculate z-score
z = (p2 - p1) / SE

# Calculate p-value (two-tailed)
p_value = 2 * (1 - stats.norm.cdf(abs(z)))
```

```
print(f"z-score: {z}")
print(f"p-value: {p_value}")
```

Results:
z-score: 0.5566920718191339
p-value: 0.5777378346579753

A p-value of 0.5777 was calculated suggests that there is a high probability that the observed difference (or a more extreme one) could have occurred by random chance alone, assuming that the null hypothesis is True.

Conclusion:

1. Statistically NOT Significant Difference: The p-value of 0.5777 indicates that the observed difference in conversion rates is statistically NOT significant. This means it is likely that the observed difference is due to random chance, and we can accept the null hypothesis that there is no difference between the forms.

2. Determine the Impact on Purchasing Behavior

To analyze the impact on purchasing behavior, we need to join the conversion data with account details and compare the purchasing behavior (first purchase) between the two variations.

Explanation:

- decisions, conversions, and account_details tables are joined using the relevant keys (`visitor_id`, `user_account_id`, and `account_id`).
- The number of sign-ups for each variation is counted.
- The number of users who made a first purchase is counted (based on account_id of users who did `first_purchase_value`) grouped by variation_id.
- The sum of first purchase value is calculated for each variation.
- The purchase rate is calculated as the ratio of purchasers (first purchase) to the number of sign-ups.

SQL Query Calculate purchasing behavior metrics for each variation can be found here

```
-- Calculate purchasing behavior metrics for each variation

WITH sign_up_count AS (
SELECT
	c.variation_id,
	COUNT(DISTINCT c.visitor_id) AS signup_count
FROM
	conversions c
where c.event_name = 'signUp'
group by
	c.variation_id
),

First_purchaser_count AS (
```

```
SELECT
    c.variation_id,
      COUNT(distinct CASE WHEN ad.first_purchase_value !=0 and
      ad.sign_up_method = 'manual' THEN
      ad.account_id END) AS first_purchasers,
      SUM(CASE WHEN ad.sign_up_method = 'manual' and
      ad.first_purchase_value !=0 THEN
      ad.first_purchase_value END) AS sum_first_purchase_value
FROM
      account_details ad
JOIN (
      SELECT
            distinct user_account_id,
            variation_id
      FROM
            conversions
      ) c ON ad.account_id = c.user_account_id
GROUP BY
      c.variation_id
)

SELECT
    su.variation_id,
    su.signup_count,
    pu.first_purchasers,
    pu.sum_first_purchase_value,
    (pu.first_purchasers * 1.0 / su.signup_count) AS purchase_rate
FROM
    sign_up_count su
JOIN
    First_purchaser_count pu ON su.variation_id = pu.variation_id
;
```

Results:

variation_id	sign_ups	purchasers	sum_first_purchase_value	purchase_rate
25974540423	3934	409	6815.91	0.10396
25991400310	3940	447	6550.53	0.11345

```
import scipy.stats as stats
import math

# Given data
n1 = 3940  # number of sign-ups for the original form
n2 = 3934  # number of sign-ups for the multi-step form
p1 = 0.11345 # purchase rate for the original form
p2 = 0.10396 # purchase rate for the multi-step form

# Calculate number of purchasers
x1 = n1 * p1
x2 = n2 * p2

# Calculate pooled proportion
p_pooled = (x1 + x2) / (n1 + n2)

# Calculate standard error
SE = math.sqrt(p_pooled * (1 - p_pooled) * (1/n1 + 1/n2))
```

```
# Calculate z-score
z = (p2 - p1) / SE

# Calculate p-value (two-tailed)
p_value = 2 * (1 - stats.norm.cdf(abs(z)))

print(f"z-score: {z}")
print(f"p-value: {p_value}")
```

z-score: -1.3526703138862823
p-value: 0.1761609782786917

The p-value is approximately 0.1762, which is greater than 0.05, indicating that there is a statistically NO significant difference between the purchase rates of the original form and the multi-step form.

Conclusion

The p-value suggests that the difference in purchase rates between the original form and the multi-step form is statistically NOT significant.

Recommendation

To effectively measure the impact of modifying a one-stage sign-up form to a three-stage sign-up form on conversion rates, it is recommended to collect various digital metrics. Here are the key metrics to focus on:

1. Conversion Rate Metrics

- Overall Conversion Rate: Measure the percentage of users who complete the sign-up process.
- Stage-Specific Conversion Rates: Measure the drop-off rate at each stage of the sign-up process.
    - Stage 1 to Stage 2 Conversion Rate
    - Stage 2 to Stage 3 Conversion Rate
    - Stage 3 to Completion Conversion Rate

2. Funnel Metrics

- Funnel Drop-off Rates: The percentage of users who abandon the sign-up process at each stage.
- Time to Complete Each Stage: Average time users spend on each stage of the form.

3. User Behavior Metrics

- Form Field Interaction: Track interactions with each form field to identify problematic fields.
- Metrics: Number of clicks, time spent, error rates, and correction attempts.
- Session Duration: Total time users spend on the sign-up process.

4. Engagement Metrics

- Page Views per Session: Number of pages viewed during the sign-up process.
- Bounce Rate: Percentage of users who leave the site without interacting further.

5. User Experience Metrics

- Heatmaps and Click Maps: Visual representations of user interactions at each stage of the form.
- Session Recordings: Videos of user sessions to observe navigation patterns and issues.

6. Qualitative Metrics

- User Feedback: Collect feedback through surveys or feedback forms at the end of the sign-up process.
- Usability Testing: Conduct tests with real users to gather insights into the user experience.