

A/B Test Analysis: Performance of Marketing Popups

Background

A/B test conducted for 8 days between 3rd July 2020 – 10th July 2020

Onsite customers divided into two buckets and shown one of two possible marketing popups

Each popup asked user whether they would like to ‘refer a friend’

Users given the option to click yes, click no or take no action

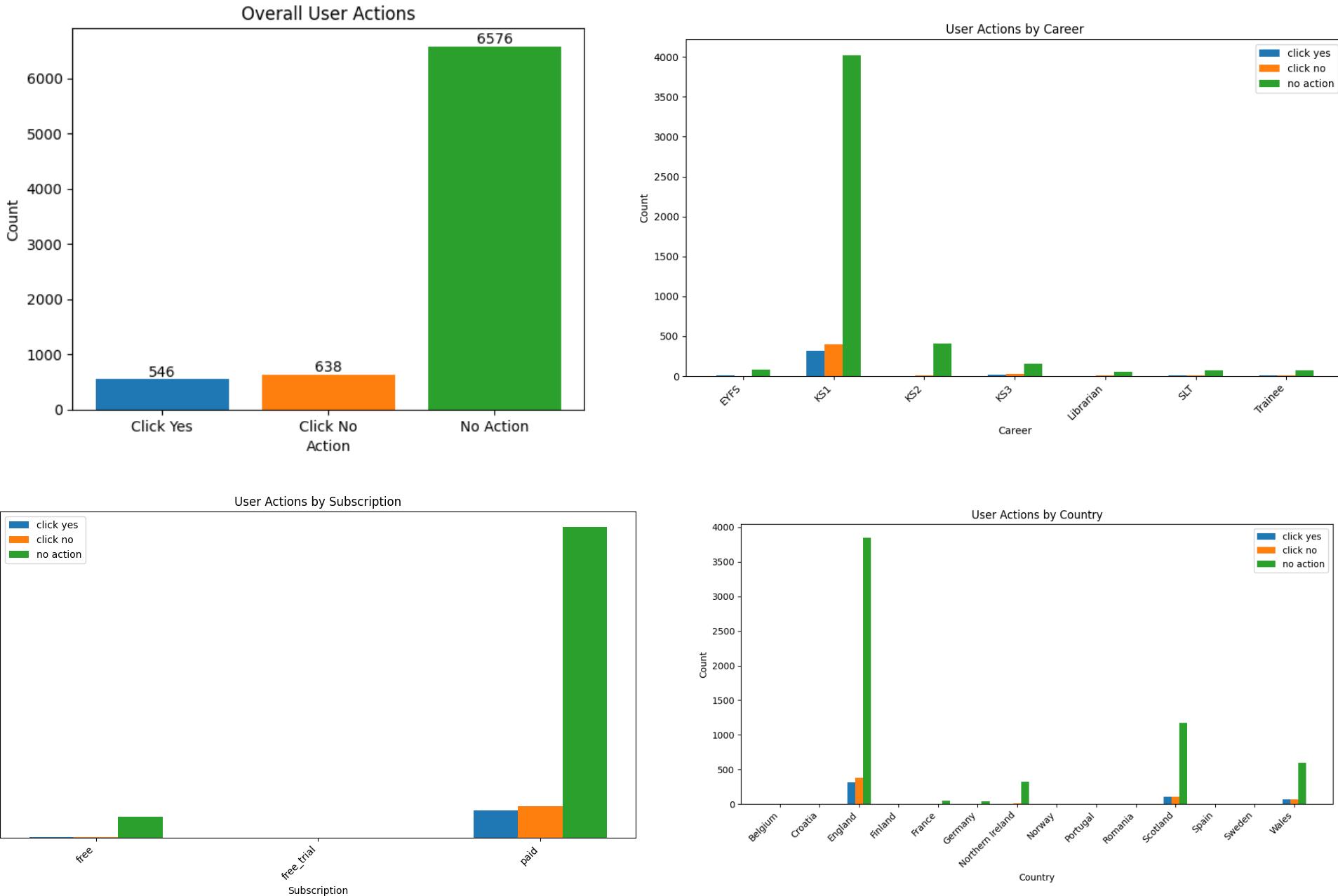


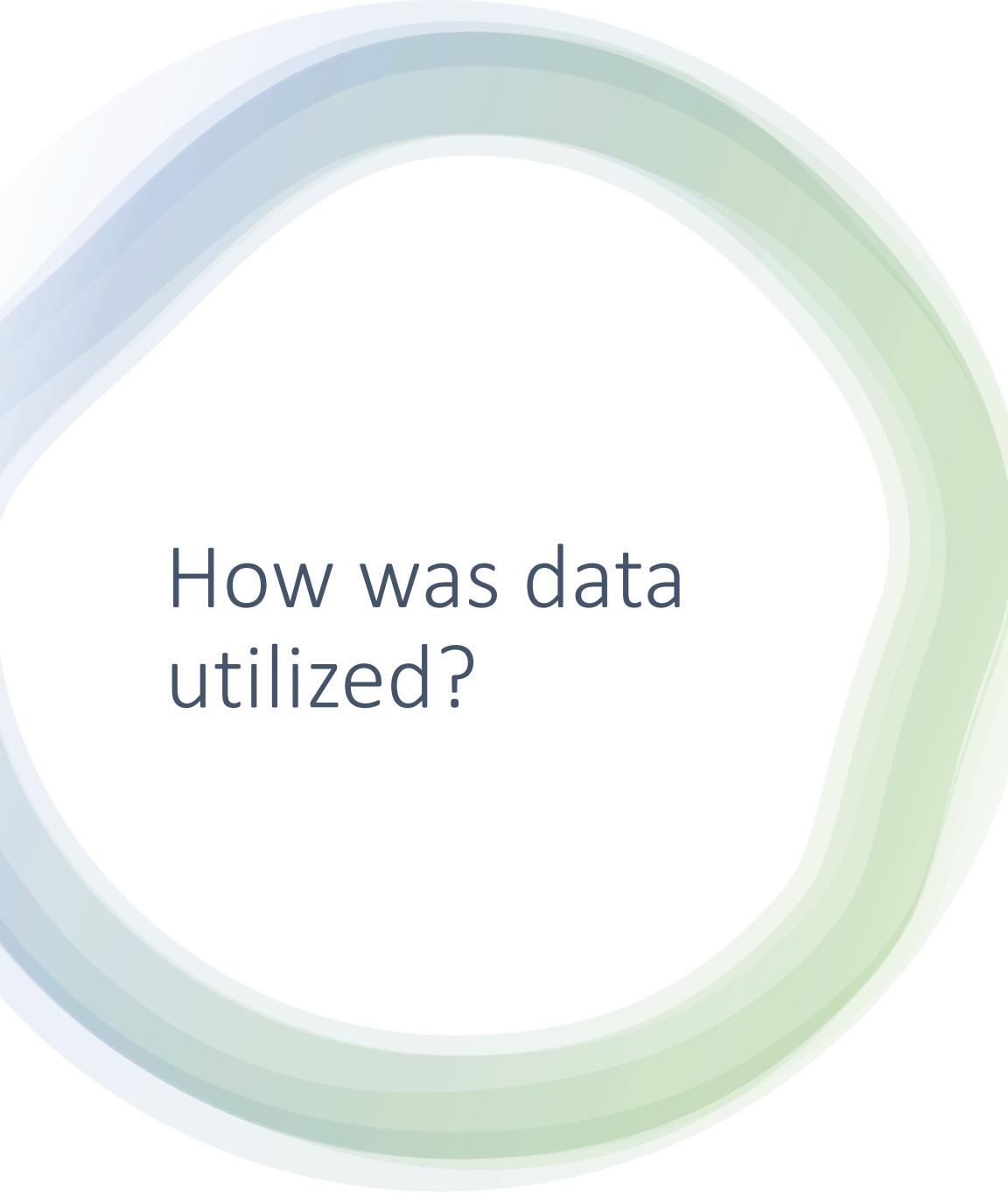
Aim

- To understand:
 - Which pop up performed better
 - Which users should be targeted with each popup

Demographics

7760 users





How was data utilized?

- Data segmented based on following factors:
 - Overall dataset
 - Country
 - Career
 - Number of searches
 - Time of day
- Missing and negative values were excluded from the analysis

How was aim addressed? 1/4

- **Statistical power**
 - Is there enough participants for an accurate result?
 - If there was not enough people it is more likely to get a false positive result

Further information on methodology

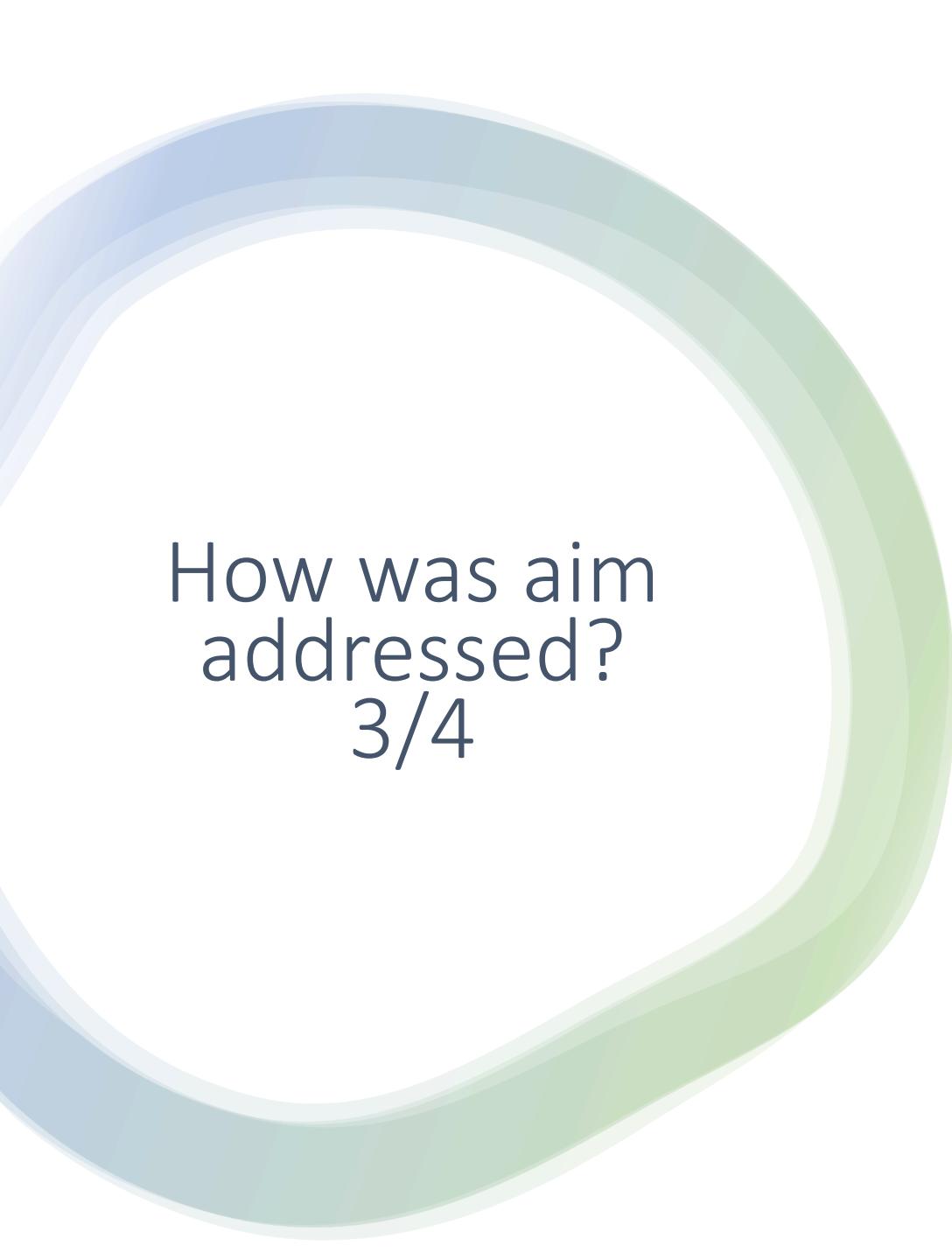
- **Statistical power**
 - Baseline conversion rate: 50-60% (chance/ slightly above chance)
 - Minimum detectable effect/lift: 10-30% (~65% conversion rate for highest performing group)
 - Statistical Significance: 5%
 - Statistical Power: 80%

How was aim addressed?

2/4

- **Chi-squared test**

- Is there a significant difference in responses? Clicking yes or no and taking action in general
- Is this effected by certain demographic e.g. country?
- 95% confidence intervals: if the test were repeated, responses to each popup would most likely be in this range of percentages



How was aim addressed? 3/4

- **Bayesian inference/analysis**
 - If there is a significant difference how significant?
 - What is the probability that the difference is due to chance?

Further information on methodology

- **Bayesian analysis**
 - Prior and likelihood function for Bayesian tests were estimated by segmenting data into two periods: first four days and last four days
 - Estimation of the prior and likelihood functions involved using a beta distribution for the priors and a binomial distribution for the likelihood function
 - The posterior distribution was inferred by combining the prior and likelihood functions, which represents updated beliefs about the parameters of interest

How was aim addressed? 4/4

- **Logistic regression**
 - How does time of day affect responses in the two groups? Is this specific to type of popup?

Further information on methodology

- **Logistic regression**
 - Logistic regression model examined the relationship between the predictors (group and time of day) and the probability of a positive outcome (taking action/clicking yes or no)
 - The main effects of group and time of day were investigated, as well as their interaction effect on the log-odds (or probability) of a positive outcome.

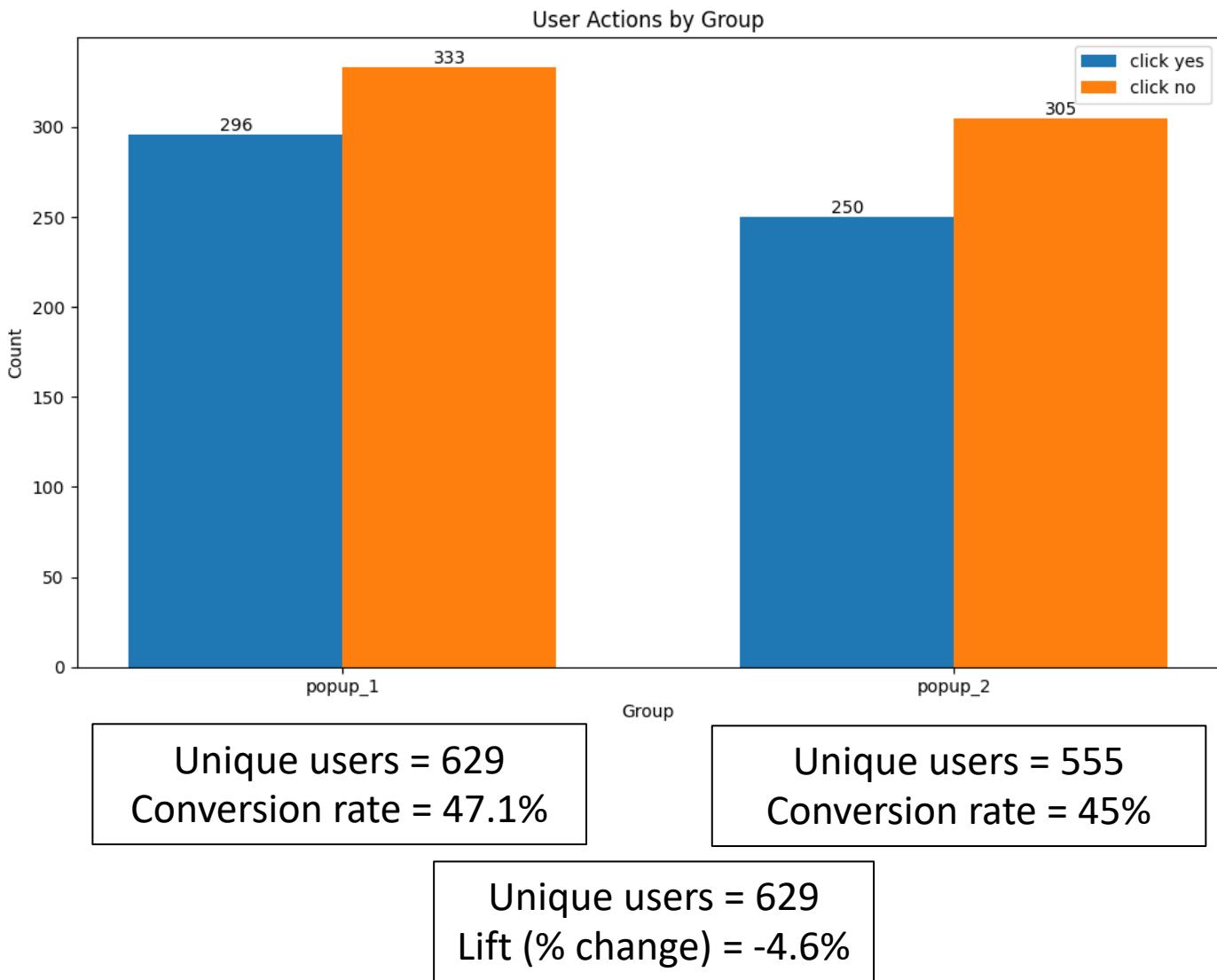


Hypotheses

- **Test 1:**
 - Clicking yes or no
- **Test 2:**
 - Taking any action vs no action
- Null hypothesis: there is no difference in conversion rate between popup 1 and 2
- Alternative hypothesis: there is a difference in conversion rate between popup 1 and 2

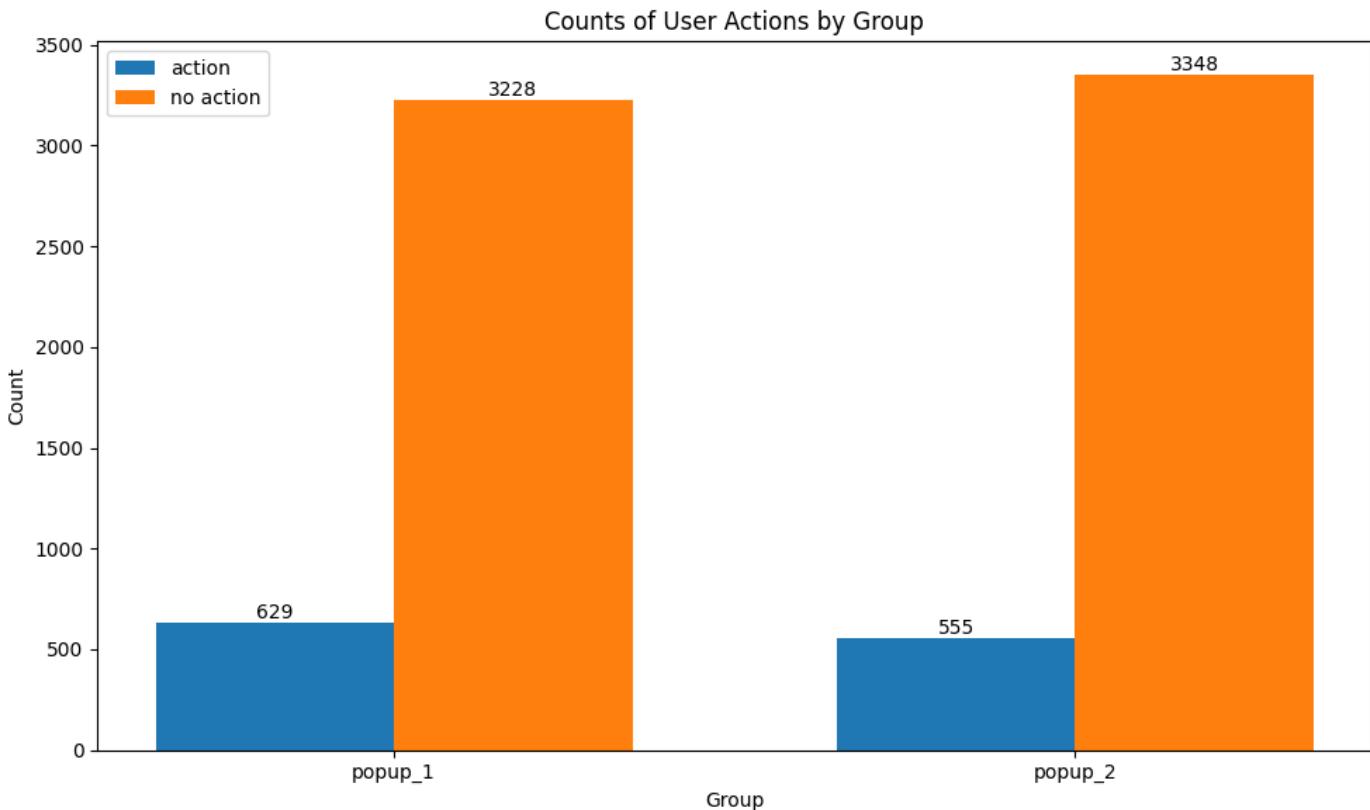
Overview of Results: Test 1

- 10.8% statistical power: high chance of false positive/negative
- Conversion rates across all demographics: 44% to 50% (popup 1) and 43% to 51% (popup 2)
- Results of all Chi-Squared tests (including demographics): non-significant
- Bayesian inference not performed as non-significant results



Overview of Results: Test 2

- 73% Statistical power: acceptable
- 95% confidence interval:
 - 15.2% - 17.5% (popup 1)
 - 13.1% - 15.4% (popup 2)
- Results of all Chi-Squared tests (including demographics): significant but with very small effect size
- Bayesian inference found around 99.6% chance that popup 1 led to more clicks



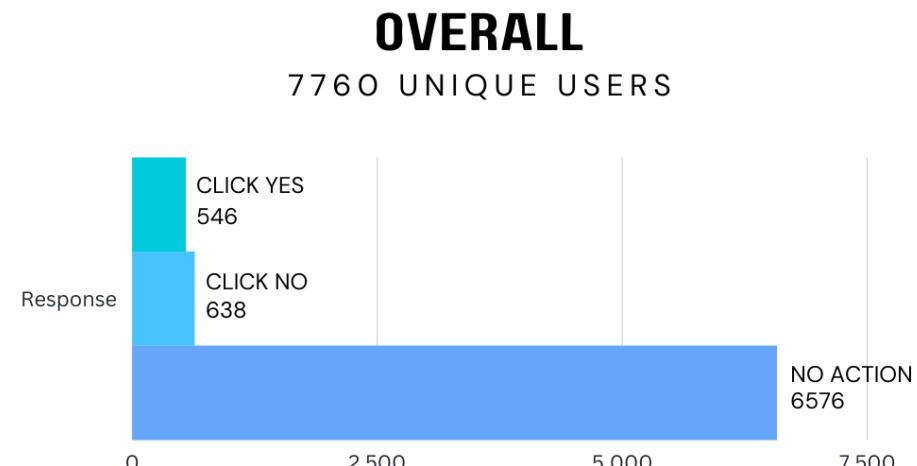
Unique users = 3857
Conversion rate = 16.3%

Unique users = 3903
Conversion rate = 14.2%

Unique users = 7760
Lift (% change) = -2.1%

Demographics

Test 1 non-significant for England users

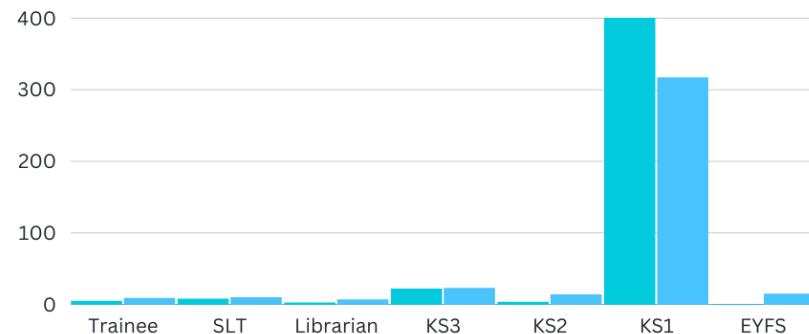


Test 2 significant for England users

Samples sizes too small for other countries

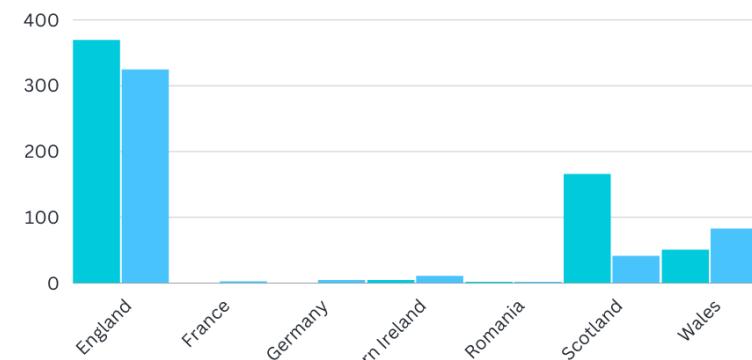
CAREERS

838 UNIQUE USERS



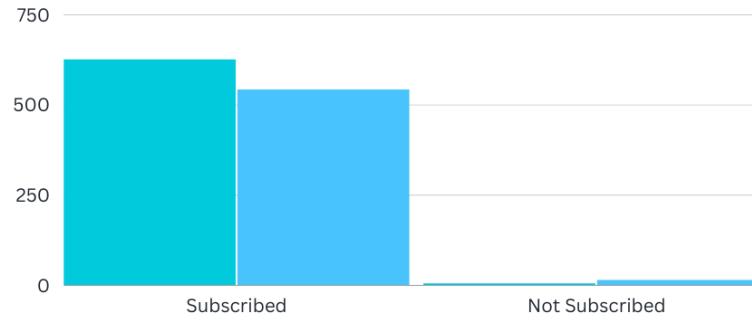
COUNTRIES

1051 UNIQUE USERS



SUBSCRIBERS

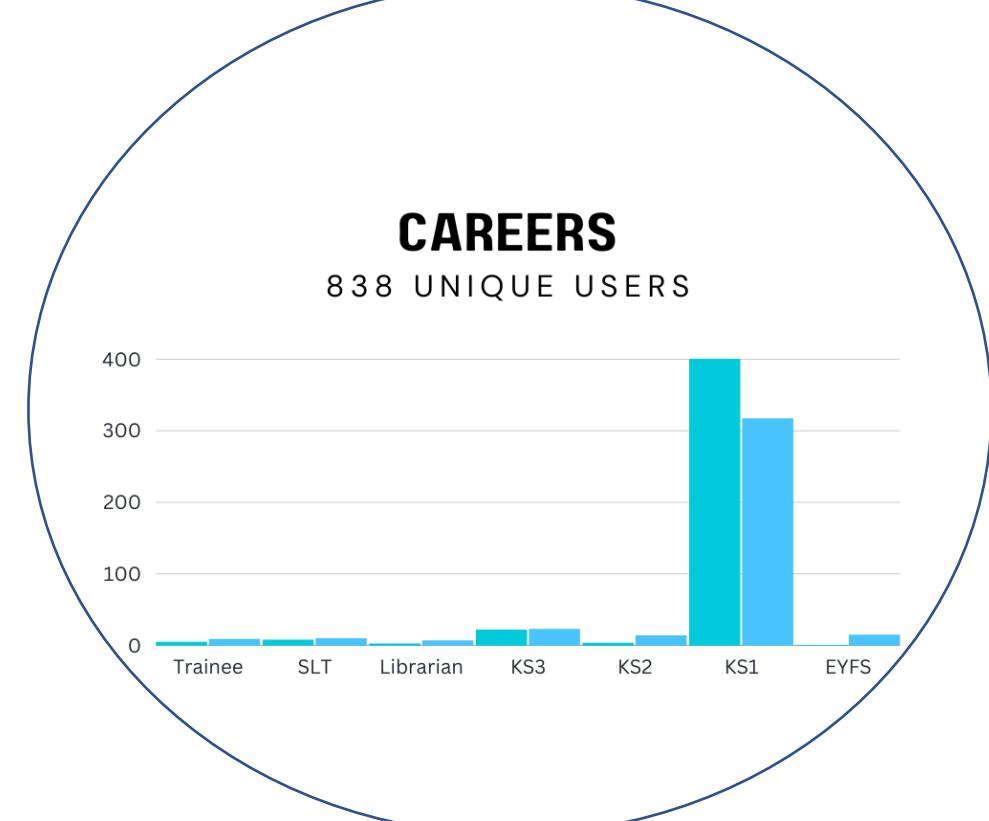
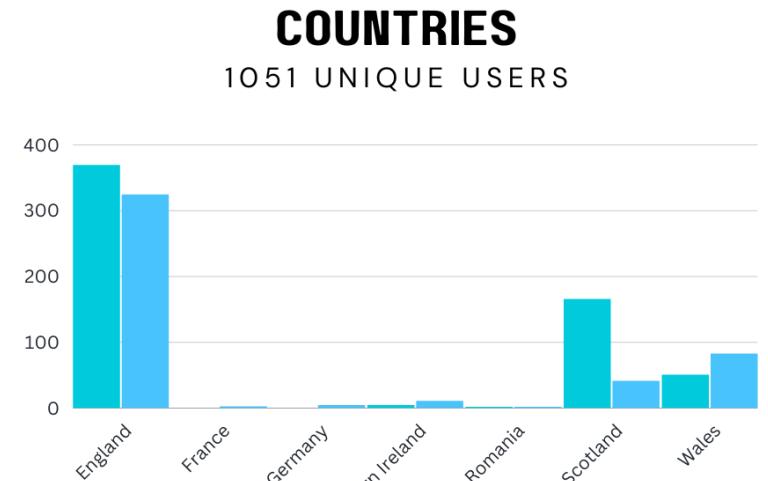
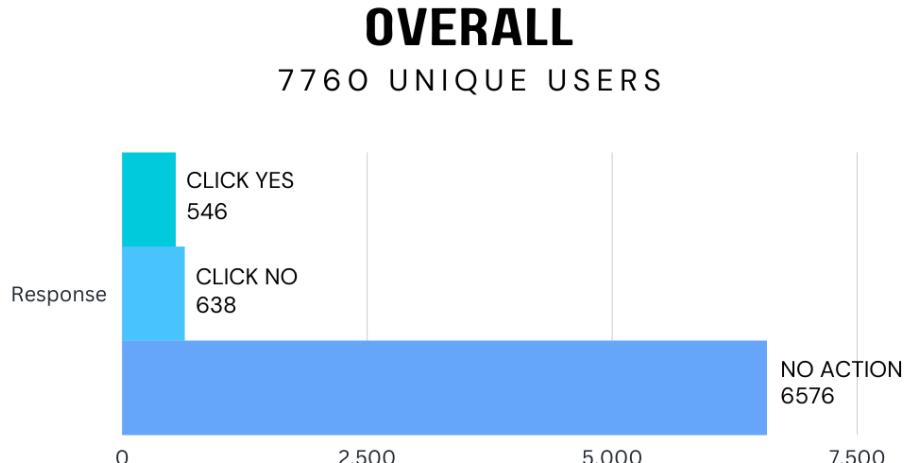
1184 UNIQUE USERS



Demographics

Same results found for
KS1

Samples sizes too small
for other countries

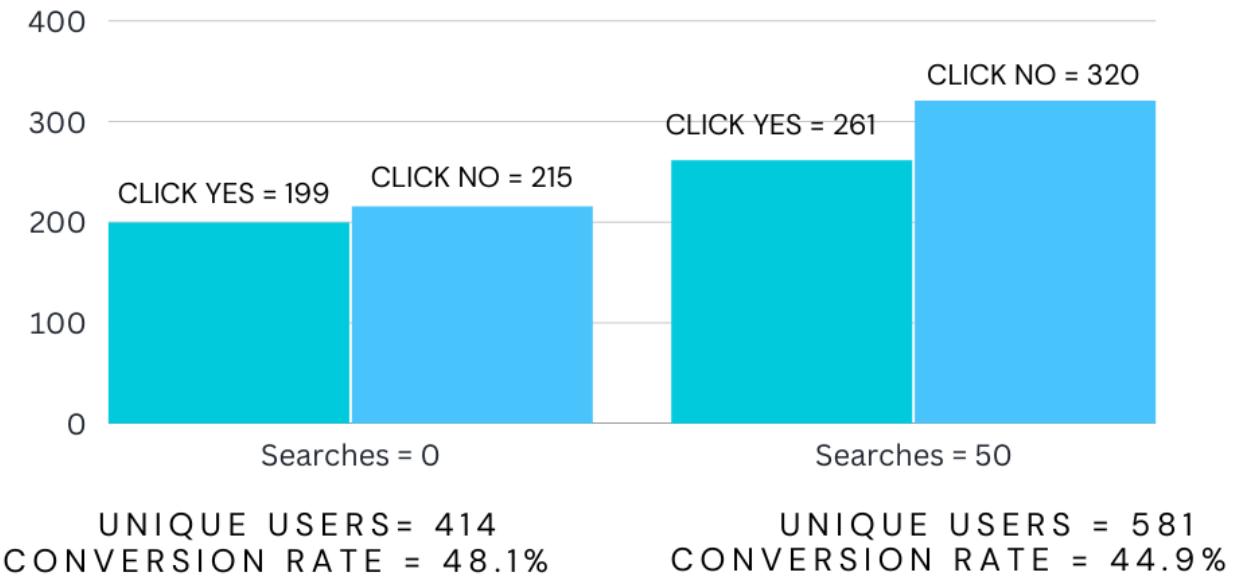


Overview of Results

– Searches

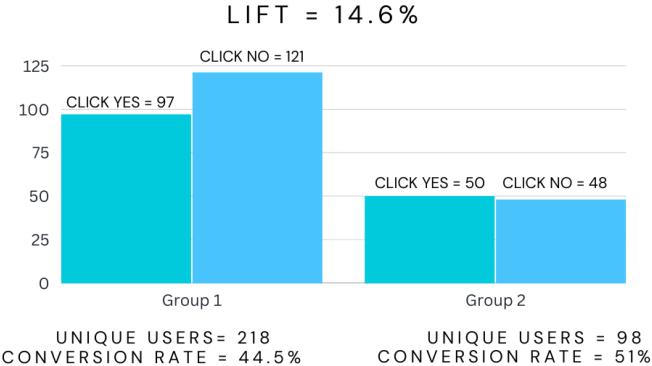
Chi-squared test for both test 1 & test 2:
non significant

SEARCHES = 0 & 50
LIFT = - 6.7%

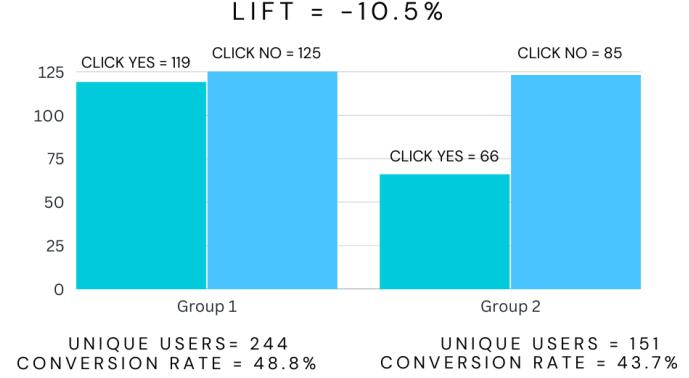


Overview of Results: Test 1 – Time of Day

TIME OF DAY - MORNING



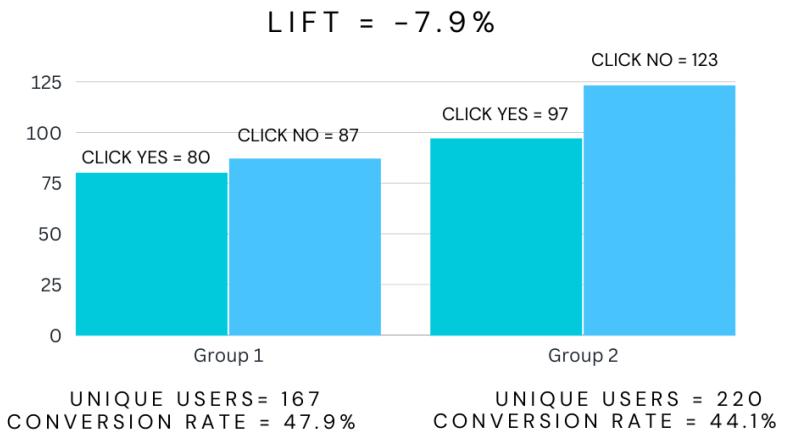
TIME OF DAY - NIGHT



Logistic regression results for interaction between time of day and group: non significant

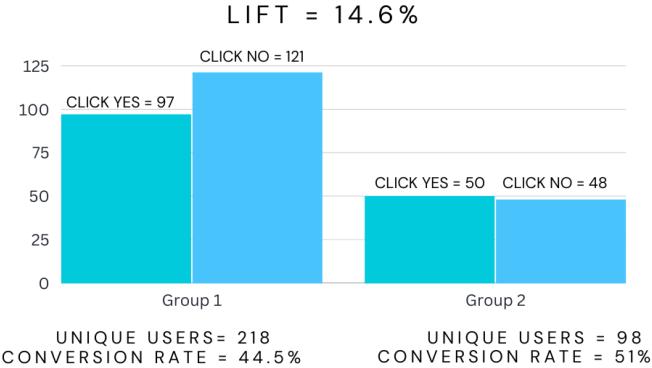
Logistic regression results for time of day and group individually: non significant

TIME OF DAY - AFTERNOON

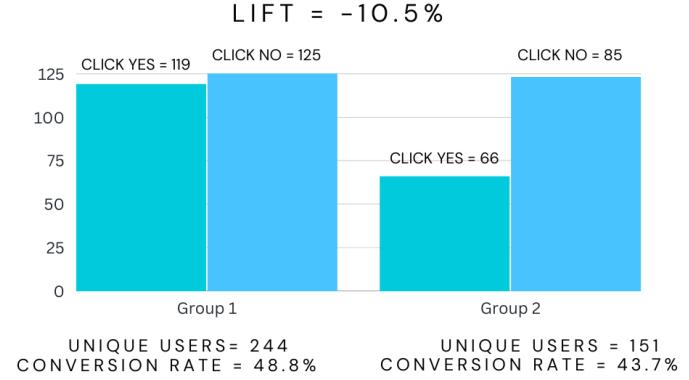


Overview of Results: Test 2 – Time of Day

TIME OF DAY - MORNING



TIME OF DAY - NIGHT

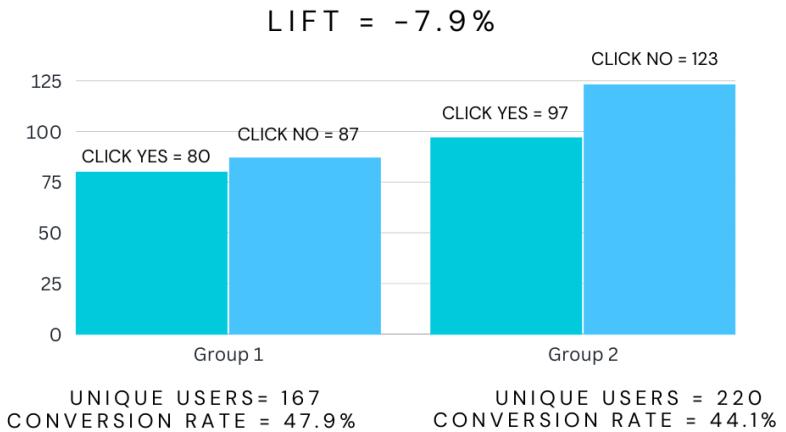


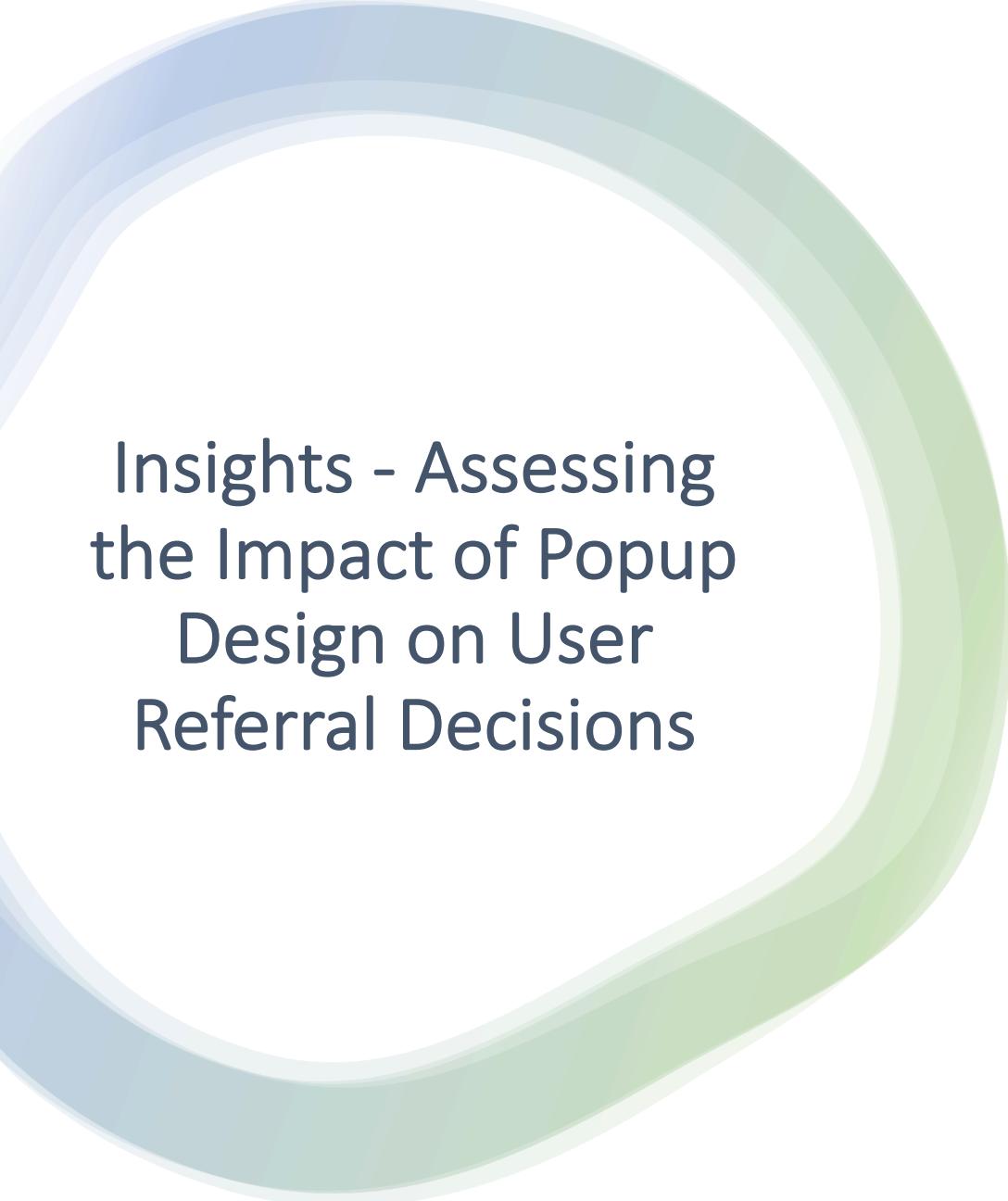
Logistic regression results for interaction between time of day and group: non significant

Logistic regression results for time of day and group individually: non significant for time of day, significant for group

The probability of taking action for a popup decreases by approximately 15.4% when moving from group 1 to group 2, holding all other variables constant.

TIME OF DAY - AFTERNOON





Insights - Assessing the Impact of Popup Design on User Referral Decisions

- Main findings:
 - Conversion rate **around chance for clicking yes or no (random)**
 - Design of popup did not influence users decision to refer a friend
 - **Design of popup had attention grabbing effect (popup1)**
 - Effect size small (**approx. 2%**)
 - Bayesian inference & logistic regression also confirm this is likely the case
 - Time of day did not influence attention to this popup or decision to refer a friend
- **However, if the effect is smaller than expected, statistical tests would need more participants to avoid false positives**



Next Steps

- Best choice would be popup 1 for attention grabbing effect
 - However, small effect
 - Possible that popup design does not influence users or **different designs should be considered**
 - Aspects of this popup could be integrated into web/app design
- More participants needed to confirm whether popup design influences users to refer a friend
- However, if popup design has **small effect on users attention**, it is **unlikely that it would influence their opinion or decision making**
- To detect small effects a substantially larger sample size for each group would be needed (for users clicking yes or no)
 - Around 9700 per group needed, 19400 in total (for 2% difference, 45% and 47%)
 - Around 1184 users clicked yes or no over 8 days so it would take around **4-5 months to collect enough data**
- **Further questions:**
 - Do popups effect user experience? i.e. number of downloads, decision to renew subscription

Overview of Results: Test 1 - England

Chi-squared test results:

Test statistic: 0.12

p-value: 0.72

Effect size: 0.02

95% confidence intervals:

Click Yes - Group 1: 0.4089858 - 0.5130588

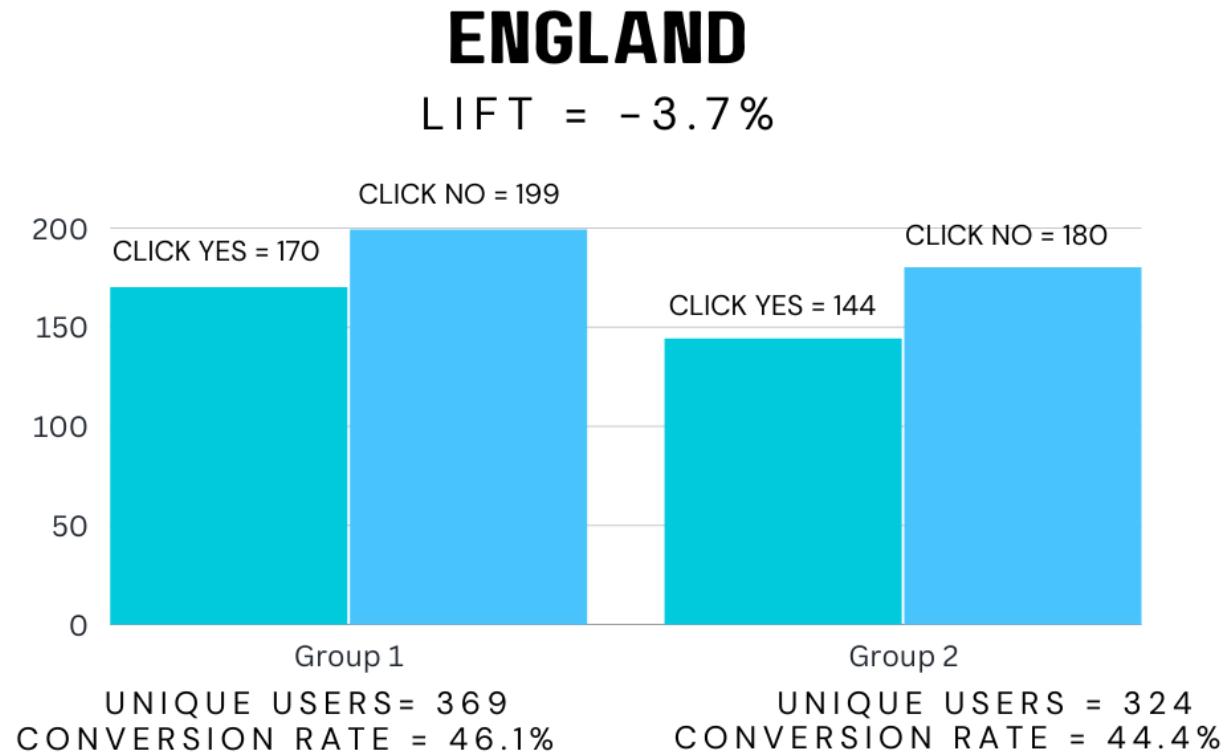
Click Yes - Group 2: 0.3895214 - 0.5003924

Click No - Group 1: 0.4869412 - 0.5910142

Click No - Group 2: 0.4996076 - 0.6104786

Bayesian test:

Probability that Group 1 had higher conversion than
Group 2: 66.7%



Overview of Results: Test 2 - England

Chi-squared test results:

Test Statistic: 12.30772
p-value: 0.0004510897

Effect size: 0.05265707

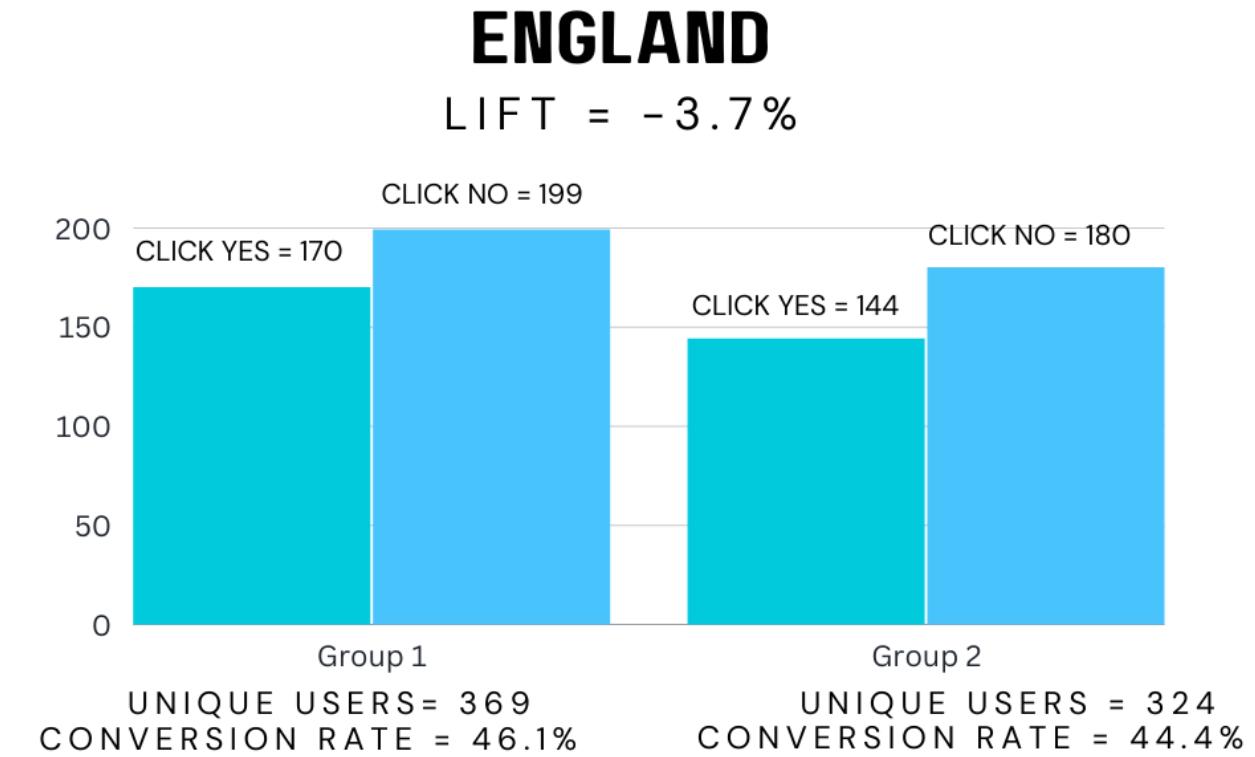
95% confidence intervals:

Action - Group 1: 0.1567963 - 0.189289

Action - Group 2: 0.1212664 - 0.1489564

No Action - Group 1: 0.810711 - 0.8432037

No Action - Group 2: 0.8510436 - 0.8787336



Bayesian test:

Probability that Group 1 had higher conversion than Group 2: 66.7%

Overview of Results – KS1

Chi-squared test results:

Test statistic: 0.13
p-value: 0.71

Effect size: 0.02

95% confidence intervals

Click Yes - Group 1: 0.4054408 - 0.5052298
Click Yes - Group 2: 0.3830862 - 0.4950457

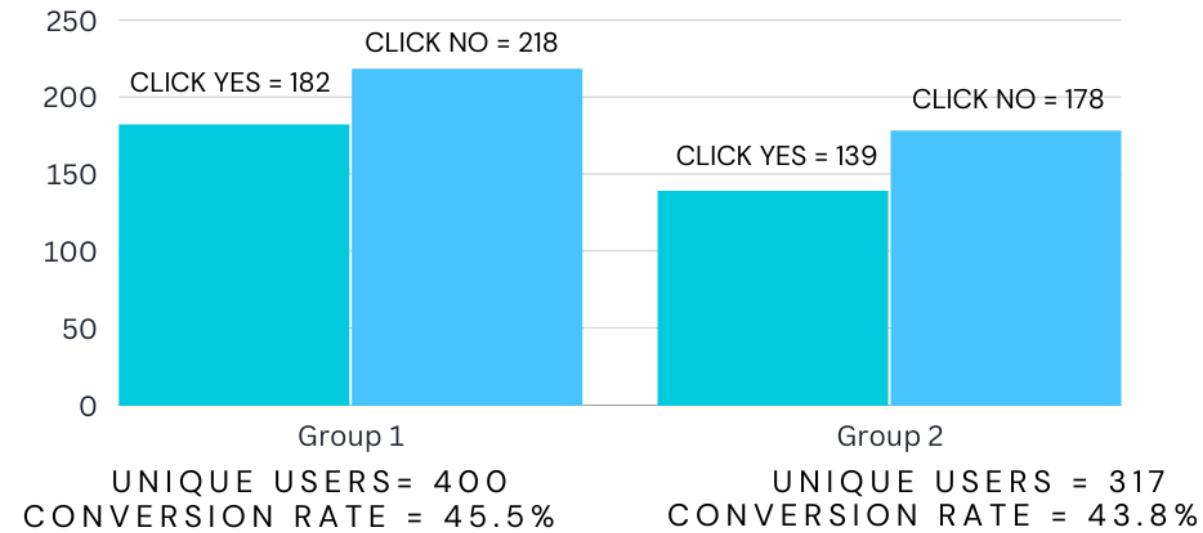
Click No - Group 1: 0.4947702 - 0.5945592
Click No - Group 2: 0.5049543 - 0.6169138

Bayesian test:

Probability that Group 1 had higher conversion than Group 2: 67.4%

KS1

LIFT = -3.7%



Overview of Results – KS1

Chi-squared test results:

Test Statistic: 14.61181

p-value: 0.0001320845

Effect size: 0.05615214

95% confidence intervals

Action - Group 1: 0.1568393 - 0.1879297

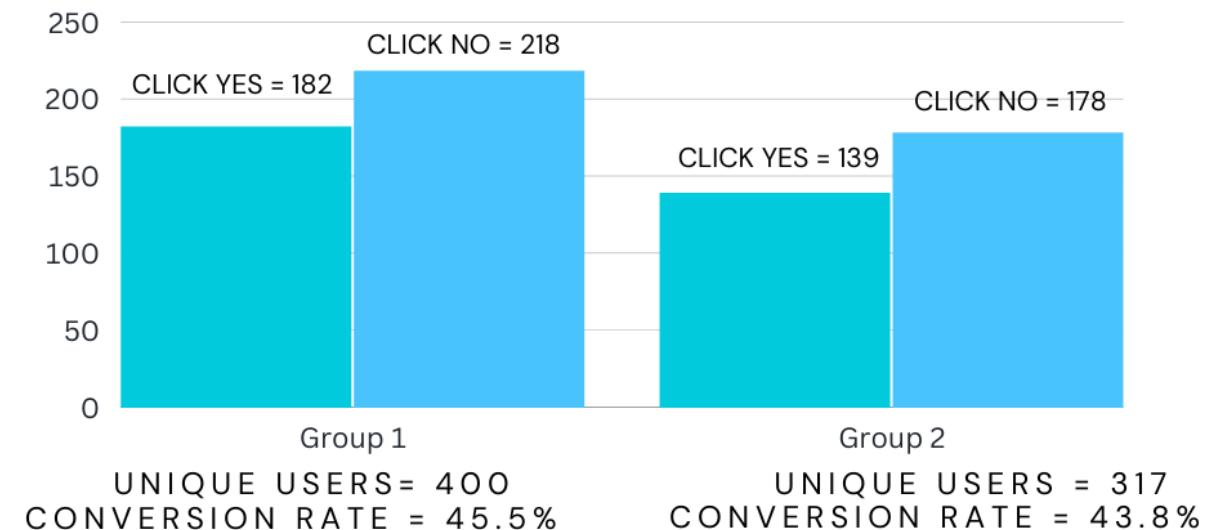
Action - Group 2: 0.1184363 - 0.1458652

No Action - Group 1: 0.8120703 - 0.8431607

No Action - Group 2: 0.8541348 - 0.8815637

KS1

LIFT = -3.7%



Overview of Results

– Searches

Chi-squared test results:

Test statistic: 0.84

p-value: 0.36

Effect size: 0.03

95% confidence intervals:

Click Yes - Group 1: 0.4316289 - 0.5300019

Click Yes - Group 2: 0.4082659 - 0.4907038

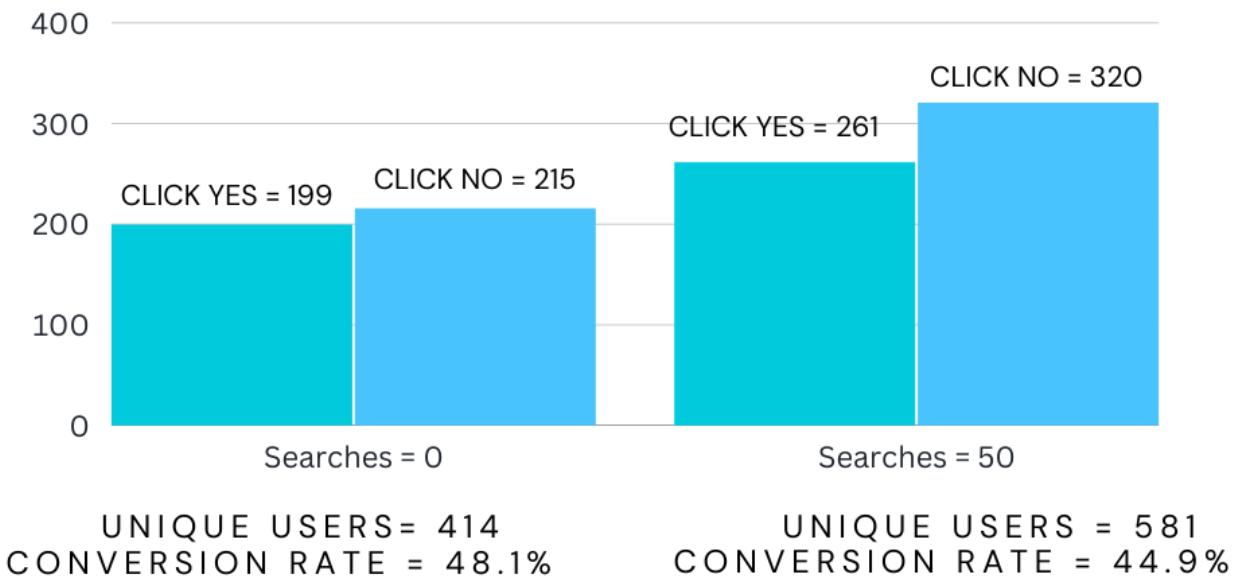
Click No - Group 1: 0.4699981 - 0.5683711

Click No - Group 2: 0.5092962 - 0.5917341

Bayesian test:

Probability that Group 0 is better than Group 50: 83.70%

SEARCHES = 0 & 50
LIFT = - 6.7%



Overview of Results: Test 2 – Searches

Chi-squared test results:

Test Statistic: 6.400908

p-value: 0.0114062

Effect size: 0.03

95% confidence intervals:

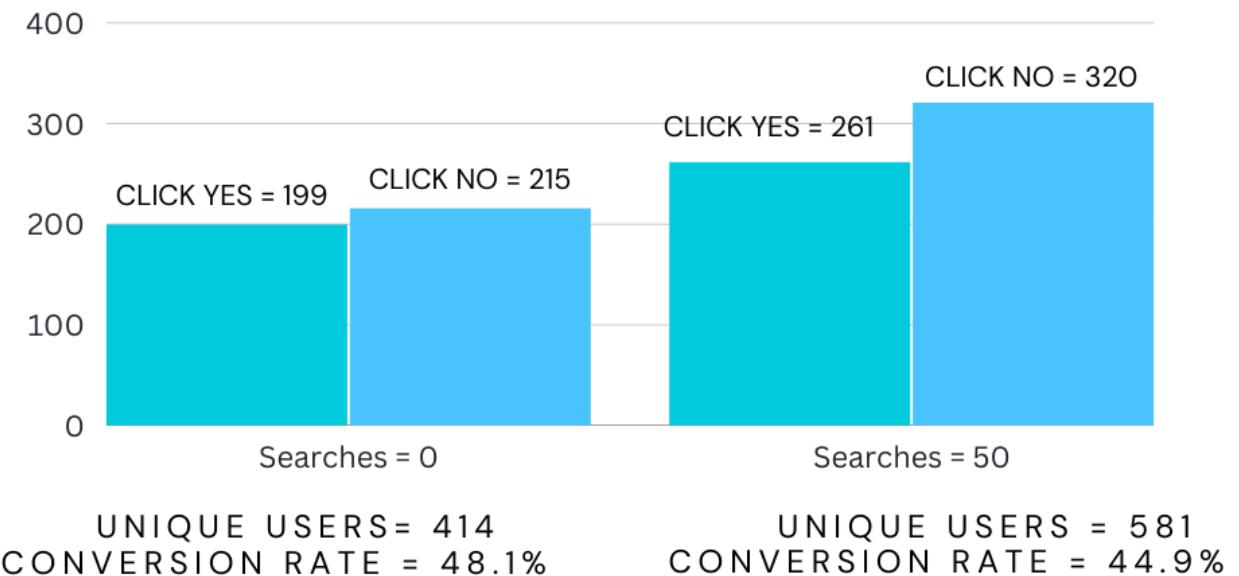
Action - Group 1: 0.1517779 - 0.181408

Action - Group 2: 0.1321738 - 0.1539157

No Action - Group 1: 0.818592 - 0.8482221

No Action - Group 2: 0.8460843 - 0.8678262

SEARCHES = 0 & 50
LIFT = - 6.7%



Overview of Results

– Time of Day

Logistic regression results::

- Time of Day:
 - Morning: The estimated coefficient for the Morning variable is 0.01553 with p-value 0.920 - represents the log-odds ratio between the morning and the reference level afternoon.
 - Night: The estimated coefficient for the Night variable is 0.03271 with p-value 0.822

Overview of Results

– Time of Day

Logistic regression results:.

- Intercept:
 - intercept value is -0.07405 with p-value of 0.738 - represents the log-odds of the outcome (clicking yes)
 - group: coefficient for the group variable is -0.06181 with p-value 0.622 - represents the log-odds ratio between the groups.

