Humans vs Nature:

Spring 2021 - W241 Project Report

Eddie Salinas, Parker Williams, Vasudev Killada

Abstract

In Travel and Leisure marketing contexts images of exciting or beautiful locations are often presented to entice customers to visit various locales. Previous research has suggested that nature images encourage interest. In the recent COVID pandemic, potential customers might be concerned about visiting destinations if they are crowded for fear of contracting or spreading the virus. We hypothesize therefore that potential tourists might want experiences with relatively few tourists to continue to practice social distancing. On the other hand, people, perhaps having been “in lockdown” might have some interest in human interaction. To explore the effects of people in images versus no people in images we endeavor to run a randomized experiment where sets of images are shown to study participants. Participants in the control group observe images without people and participants in the treatment group observe images with people. To compare the two cohorts an online advertising infrastructure is used to compare click-through rates of the two groups. If concerns to avoid people overrides interest in the locale, then images without people may have higher click-through rates ; but if interest in social interaction outweighs fears of COVID, then click-through rates of images with people would be higher.

Background

Nature images have been shown to elicit higher memory scores in both unaided recall and recognition compared to identical advertisements displaying a variety of other attractive pictures.[1] We want to know if adding humans to the nature images enhances performance of travel advertisements, or dilutes the effect of nature images.

Via Facebook, we will set up two identical advertising campaigns, and use the same targeting parameters. We will ensure that there is no overlap in people between the two groups. Through this framework we will be able to randomly assign treatment and control. The images used in the experiment will be identical in all aspects aside from whether or not there is a person or persons in the image. We will then be able to compare click through rates between the two groups to estimate our average treatment effect.

Research Question & Hypothesis

Do nature images targeted at travel enthusiasts on Facebook have higher click through rates with or without people in them?

A study by Hartmann et al.[1] showed that nature imagery elicits positive reactions; we believe in turn that such images would serve as a substrate to promote participant engagement and interaction. In that context, offering images without and with people as the treatment, we hope to better understand the psychology of travel during the COVID19 pandemic. It is our hypothesis that images without people would generate a higher click rate, but these experiments will provide data to support or perhaps refute this null hypothesis.

Experimental Design & Details

Experimental Overview

We plan to use the facebook ad platform and recruit social media users as part of this experiment. With a budget of 500 hundred dollars over a month's time period, we will capture the click rate/counts between the two groups. The Facebook ad platform provides a framework where the users are not aware of the grouping categories that we have designed behind the scenes. The only administration treatment is the type of the image that will be shown. The number of clicks and impressions is what we capture after the administration of the ad campaign clicks. We have ensured during setup that individuals in treatment cannot be in control and vice-versa.

Since we are using the Facebook platform for this experiment we are limited in the data that we can capture. We are unable to observe demographic characteristics of individuals in the treatment and control groups to conduct a balance test. To protect user privacy, Facebook shares aggregated data. The data we do have access to is the amount of money we spend, the number of impressions (not unique), number of unique individuals that have seen the ad, and the number of clicks (not unique). With these aggregated data we can calculate clickthrough rates and compare the control images with no people to the treatment images that contain people.

We utilize Facebook's A/B experiment platform to ensure that there is no cross-contamination with individuals in treatment receiving control images and vice-versa. Since we have no way to see individual data for privacy reasons as stated earlier, we do not have a means of confirming there are no spillovers or cross-contamination. We trust Facebook's experimental process works as it is intended to.

Administration of Treatment

After utilizing Facebook's built in power estimators as well as our own power calculations explained in a future section, we decided to administer the treatment across 4 different nature landscapes: desert (arches), mountain lake (dock), rainforest, and beach. An additional factor was the prominence of individuals within images. In the mountain lake and desert arches photos humans are prominently displayed in the center of the image and account for approximately 10% of the image's total pixels. In contrast the beach and rainforest images have more subdued human presence that take up less than 10% of the image's total pixels.

Desert Arches

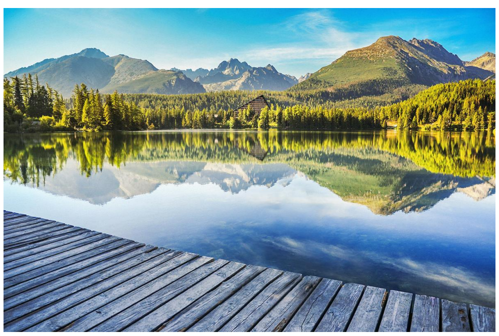
Control

Treatment

Human prominently featured and ~10% of pixels

Mountain Lake

Control

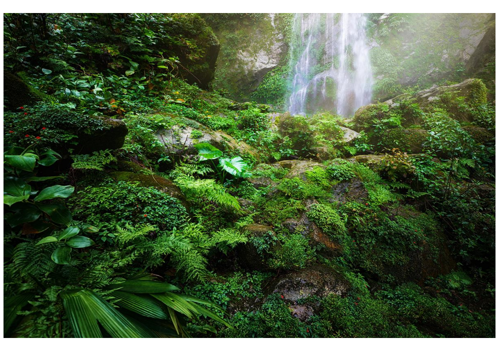


Treatment

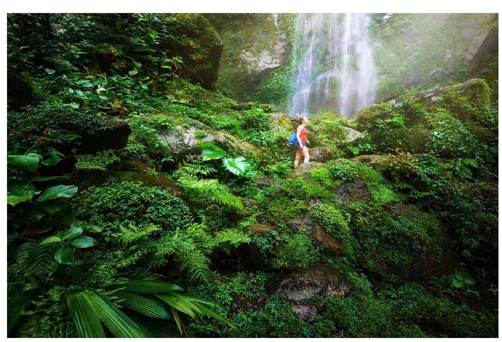
Human prominently featured and ~10% of pixels

Rainforest

Control



Treatment



Subdued human presence <10% of image pixels

Beach

Control Treatment

Subdued human presence <10% of image pixels

Power Calculations

Statistical power, in the context of this project, is the probability of finding that two click-rates are different when they are in fact different. Two-tailed power was computed using sampling distribution of proportions calculations. The sampling distribution of proportions is normally distributed mean-centered (at the sampled value) but with a standard deviation of $\sigma_{\rho} = \sqrt{\frac{p(1-p}{n}}$. To avoid situations where a normal distribution might not work, for example if a truncated normal distribution were necessary, we only considered scenarios where *np>=10* and *n(p-1)>=10*. In all cases, a "critical value" was computed from the control's distribution, using R's[2] "qnorm", for the two-tailed power at the $1 - \frac{\alpha}{2}$ quantile. From this critical value, $\boldsymbol{\beta}$(power) was computed using the cumulative distribution function("pnorm") from the competing treatment sampling distribution of proportions, and finally a *power* as $1 - \beta$.

Before our experiment began, based on our $500 budget, prior experience, and online review of others' experience with FaceBook ads[3], we estimated that we might observe click rates around 0.001 up to 0.01 (0.1 percent and 1 percent respectively). Given our budget, and estimated ranges of click rates, using the formula

$$\sigma = ic_i + ic_c\mu \tag{1}$$

we were able to infer a range of impressions. Here σ is the total cost, i the number of impressions, c_i and c_c the cost-per-impression, and cost-per-click, and μ the click-rate. The formula yielded impressions from *5,000* up to *30,000*. Estimates of cost-per-click and cost-per-impression were also gleaned from others' experience with facebook ads[3]. We note too that Facebook provided to us an estimated power of *0.66* unconditional of which photos were being used. We speculate this value may be an internal prior value. Facebook's power estimates only marginally increased when adding more spend to the campaigns. A campaign only testing 2 images using the full 500 dollars only had around *.70* power according to the Facebook-provided estimate. Given our limited budget, we made the trade-off going with a lower power estimate of *.66* compared to *.70* motivated by the ability to conduct more tests.

Not surprisingly, if few impressions are made or if there are few clicks and especially if the two rates are near each other, then power is low. Furthermore, all realized values were within the ranges of estimated possibilities using the aforementioned formula. Plots below summarize the power calculations using these estimated values. Soon after the ad campaigns began, initial observations of data were made ; these preliminary observations showed that rates varied between 1.4% to 2.6%. Once the experiment was completed, final power calculations were made, and none of the 4 tests were powered above *66%* (with two-tailed tests). The 4 campaigns (Arches, Beach, Dock, and Rainforest) had corresponding observed powers of *66%*, *2.6%*, *4.7%*, and *54%* respectively. The observed power values, post-experiment are displayed in the last figure below.

Power at Various Click Rates and Impressions

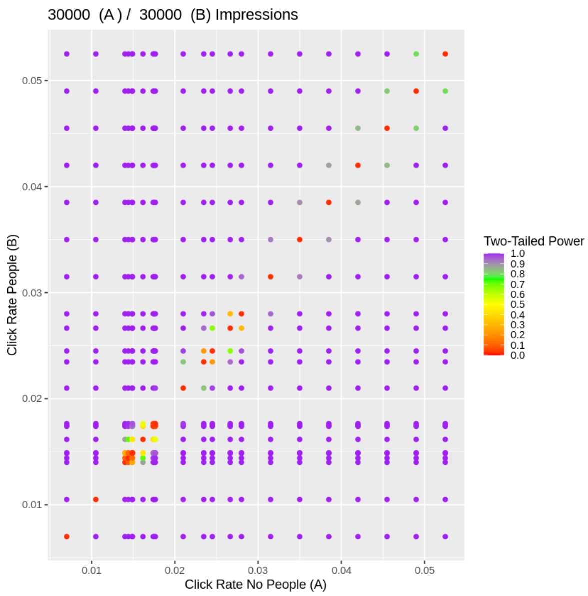


Figure PA

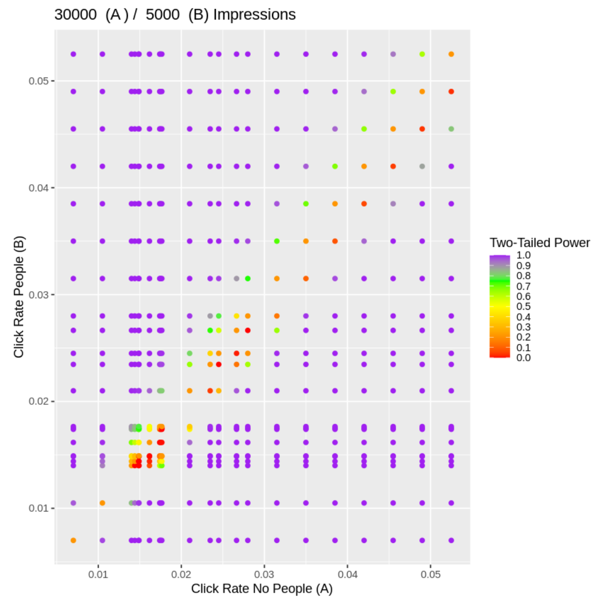


Figure PB

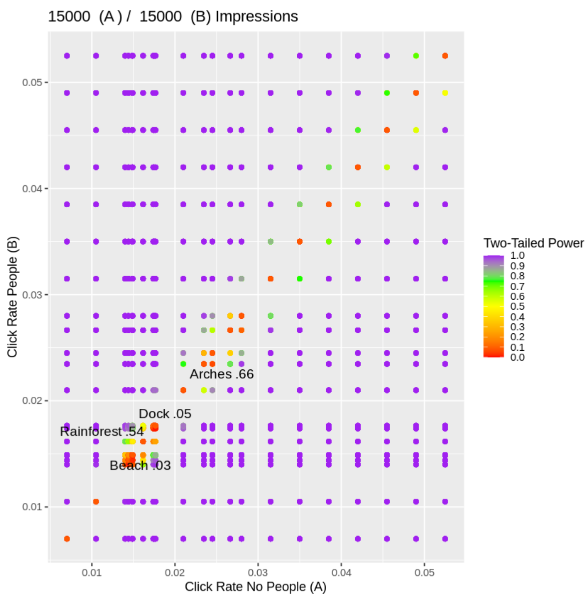


Figure PC

Figure PA : Pre-experiment power for a click-rate A/B test if both A & B have 30,000 impressions each. A sharp increase is observed off the diagonal.

Figure PB : Pre-experiment power for a click-rate A/B test if A & B have 30,000 and 5,000 impressions respectively. The asymmetric impression counts induce corresponding asymmetric increases in power off the diagonal.

Figure PC: On the 4 A/B tests, empirical impression counts were observed ranging from a low of 12,608 and a high of 19,514 with a median of 15,462. This graph groups the 4 A/B empirical tests into the existing power calculations for hypothetical tests of pairs with 15,000 impressions each. The 4 campaigns (Arches, Beach, Dock, and Rainforest) had corresponding powers of 66%, 2.6%, 4.7%, and 54% respectively which are annotated here though each of those powers were observed at the various impression counts.

Analysis & Hypothesis Testing

After and during the campaigns we downloaded the results from Facebook. These data were aggregated and not at the individual level; this supports facebook user-privacy, but also prevents methods like randomization and balance checks. For the Rainforest set of Ads we have 281 clicks and 19,514 For the Rainforest set of Ads we have 309 clicks and 19,115 For proportions, the sampling distribution may be approximated with a normal distribution whose mean is the sampled proportion, but whose standard deviation σ is equal to $\sqrt{\frac{p(1-p)}{n}}$. We have 9e-04 and 9e-04 as standard deviations for the sampling distributions for no people and people respectively. To compare the two proportions we may set $H_0: p_p = p_{np}$ that the two proportions are equal. We use the prop.test function to compute p-values in our testing. The p-value 0.1697 indicates that we fail to reject the null hypothesis that the two proportions are equal at the α=0.05 confidence level.

For the Arches set of Ads we have 336 clicks and 12,608 For the Arches set of Ads we have 307 clicks and 13081 We have 0.0014 and 0.0013 as standard deviations for the sampling distributions for no people and people respectively. The p-value 0.1115 indicates that we fail to reject the null hypothesis that the two proportions are equal at the 0.05 confidence level.

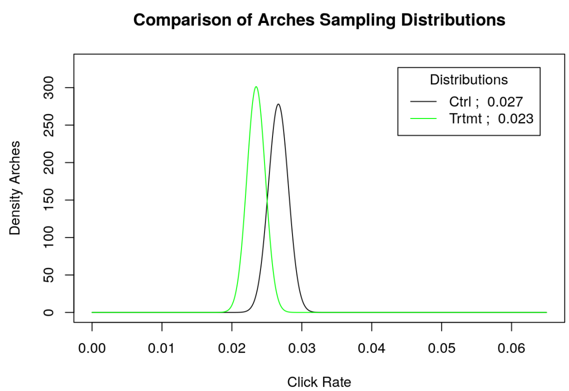
For the Dock set of Ads we have 258 clicks and 14,833 For the Dock set of Ads we have 250 clicks and 14161 We have 0.0011 and 0.0011 as standard deviations for the sampling distributions for no people and people respectively. The p-value 0.9012 indicates that we fail to reject the null hypothesis that the two proportions are equal at the 0.050 confidence level.

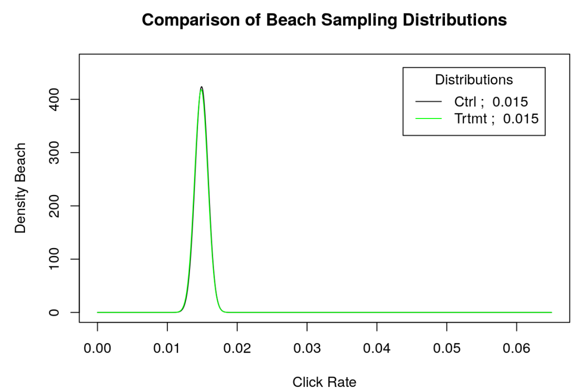
For the Beach set of Ads we have 247 clicks and 16,571 For the Beach set of Ads we have 239 clicks and 16,090 We have 9e-04 and 0.0010 as standard deviations for the sampling distributions for no people and people respectively. The p-value 1.0000 indicates that we fail to reject the null hypothesis that the two proportions are equal at the 0.05 confidence level.

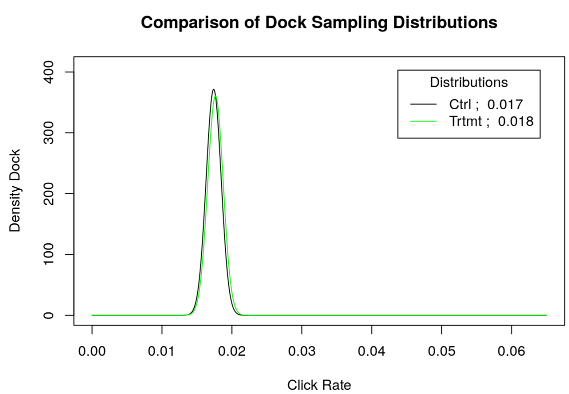
These principal data, results and analyses are summarized in the following table and figures.

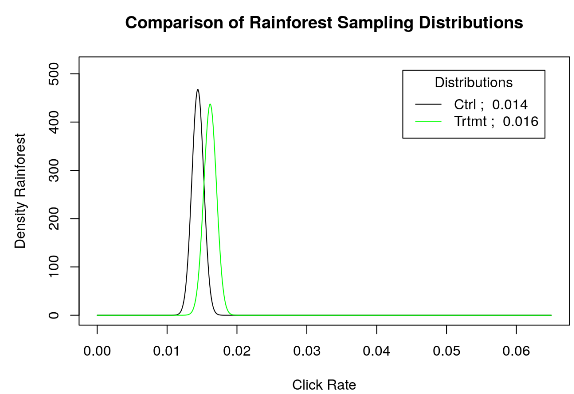
Campaign	No People (Control)			People (Treatment)			Observed Power	P Value
	Imp*.	Clicks	Click Rate	Imp.*	Clicks	Click Rate		
Arches	12,608	336	0.027	13,081	307	0.023	0.659	0.112
Beach	16,571	247	0.015	16,090	239	0.015	0.027	1.00
Dock	14,833	258	0.017	14,161	250	0.018	0.048	0.901
Rainforest	19,514	281	0.014	19,115	309	0.016	0.541	0.170

*Imp meaning "impressions". Figures in green are closest to statistical significance.









The densities above are the sampling distributions of proportions; they are computed as approximations to a normal distribution whose mean is the empirical mean, but whose standard deviation is $\sqrt{\frac{p(1-p)}{n}}$ where n and p are the number of impressions and the corresponding proportion mean.

To further explore the data and any impacts of people in images on click rates, we pooled the data and defined five regression models. The models:

1. *People ,*
2. *People and Scene,*
3. *People (interacting) with HighPeople(>=10% of pixels) and Scene,*
4. *People (interacting) with HighPeople, and*
5. *HighPeople*

The models' coefficients were computed with the R "lm" function and were augmented with robust standard errors by the "sandwich" package[4]. The weights are displayed in the Stargazer-generated[5] table below:

Effects of Images and People In Them On Clicking Those Images (w/Robust SE)

	Dependent variable:				
	Click				
	(1)	(2)	(3)	(4)	(5)
as.factor(Scene)Beach		-0.010***	-0.011***		
		(0.001)	(0.001)		
as.factor(Scene)Dock		-0.008***	-0.008***		
		(0.001)	(0.001)		
as.factor(Scene)Rainforest		-0.010***	-0.011***		
		(0.001)	(0.001)		
People	0.00003	-0.0001	0.001	0.001	
	(0.001)	(0.001)	(0.001)	(0.001)	
PeopleHigh				0.007***	0.006***
				(0.001)	(0.001)
People:PeopleHigh			-0.002	-0.002	
				(0.002)	
Constant	0.018***	0.025***	0.026***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0005)
Observations	125,973	125,973	125,973	125,973	125,973
R^2	0.00000	0.001	0.001	0.001	0.001
Adjusted R^2	-0.00001	0.001	0.001	0.0005	0.0005
Residual Std. Error	0.132 (df = 125971)	0.132 (df = 125968)	0.132 (df = 125967)	0.132 (df = 125969)	0.132 (df = 125971)
F Statistic	0.002 (df = 1; 125971)	26.920*** (df = 4; 125968)	22.004*** (df = 5; 125967)	21.753*** (df = 3; 125969)	63.230*** (df = 1; 125971)
Note:					*p<0.1;* ***p<0.05;*** p<0.01

Focusing on the role of people, we see that having any person in any image is never significant at any of the confidence levels (0.1, 0.05, or 0.01). This is consistent with the primary results summarized earlier with no p-value <0.1. Second though, we note in models 4 and 5 that the specific "PeopleHigh" treatment (where ten percent or more of pixels are "people") that they are significant at every confidence level. Though these are significant we refrain from declaring that "PeopleHigh" is an effective treatment to induce clicking. We refrain because not only are the R^2values low, but these are only two experiments (Dock and Arches) and we would insist on more experiments before making such a declaration. We believe that these values however hint at subsequent experiments that may provide further insight about the impacts of people in images on click rates.

Conclusion and Potential Next Steps:

Given our limited budget and relatively underpowered tests, unsurprisingly we did not obtain statistically significant results for any of our 4 tests. Interestingly the two tests that were the closest to being statistically significant have the least in common: the desert and rainforest photos show opposite geographical environments. Moreover, the prominence of people is also opposite between the two images (desert = high, rainforest = low). The mountain lake and beach photos had nearly identical results between treatment and control. Our conclusions were further supported by regression analysis. In all cases, we failed to reject the null hypotheses that including people in images changes the click-rate in a statistically significant way.

We had two major hurdles in achieving results, and they are also the places we would make changes in the future. First, we would find a way to spend more money on the problem to get more power from the start. Besides additional funding, perhaps partnering with the national parks service, who might also have interest in answering this question, might be fruitful. Second, editing photos is difficult and takes time; with more time it would be interesting to test images that show large crowds of people at specific nature locations vs no people. This additional factor would better help explore the psychology of travel during a pandemic event.

References

1. Patrick Hartmann, Vanessa Apaolaza & Patxi Alija (2013) Nature imagery in advertising, International Journal of Advertising, 32:2, 183-210, DOI: 10.2501/IJA-32-2-183-210
2. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/
3. Cost information https://www.webfx.com/social-media/how-much-does-facebook-advertising-cost.html
4. Zeileis A (2004). "Econometric Computing with HC and HAC Covariance Matrix Estimators." *Journal of Statistical Software*, **11**(10), 1–17. doi: 10.18637/jss.v011.i10.
5. Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.1. https://CRAN.R-project.org/package=stargazer