Effect of Text Format in an online AD on Intent to Purchase

W241 Final Project - Fall 2021

Kumar Narayanan, Ryan Goding, Melinda Leung, Vishakh Pillai

12/07/2021

Abstract

The AD industry, both the product owners and the AD networks, are always on the lookout to get more clicks on the ADs they place in various web sites. Online ADs drive revenues of several large corporations, and is one of the chief avenues used by the products and services industry to reach their respective potential consumers. A click on an AD by an online users is revenue for the entire AD ecosystem, while, at the same time, it allows the product or services owning company gauge the level of interest, which influences the potential sale they can achieve. The AD industry is constantly varying various elements of an AD to at least get a click, if not getting the customers make a purchase.

We wanted to understand the impact the formatting of letters in an AD will have on its ability to attract attention from visitors to websites such as a marketplace or selling platforms. The approach we took was to create two ADs for a used car that only differs in the formatting of the wordings. In one of the ADs, the **Treatment AD**, all the words that describe various aspects of the car have been fully capitalized, while the **Control AD** had words the normal way - i.e only the first letter is capitalized, as in any normal writing.

The two ADs side-by-side illustrate what we advertised. This is one of the eight images we show below. All 8 images that formed the Treatment Group of ADs had the description all capitalized while those in the control group had normal lettering.

2009 ACURA TL - $7,500	2009 Acura TL - $7,500
image 1 of 8	image 1 of 8
ACURA 2009 TL, SEDAN 4 DOOR, MILEAGE: 167,584, NON-RENTAL CAR, ONE PREVIOUS OWNER, V6, VTEC, 3.7 LITER, SH-AWD, HILL ASCENT CONTROL, POWER STEERING, LEATHER SEATS, NAVIGATION SYSTEM, AIR CONDITIONING, HEATED SEATS, ANTI-THEFT SYSTEM, GREAT CONDITION WITH MINOR WEAR, BACKUP CAMERA, MOONROOF	Acura 2009 TL, Sedan 4 Door, Mileage: 167,584, Non-Rental car, One Previous owner, V6, VTEC, 3.7 Liter, SH-AWD, Hill Ascent Control, Power Steering, Leather Seats, Navigation System, Air Conditioning, Heated Seats, Anti-Theft System, Great Condition with Minor Wear, Backup Camera, Moonroof
Treatment	**Control**

Background

With the ever increasing options of goods one can purchase in online marketplaces, sellers need to identify more ways to capture the attention of their audience. Experimenting with fonts and colors is one method to do so — it is heavily leveraged by marketers as a means to elicit emotion, promote interest, and drive sales. This theory has already been tested on physical packaged foods. *Waheed et al.* (2018) discovered that product packaging, color, font style and overall design has a significant impact on the consumer purchase intentions.

We wanted to test this effect on the growing online marketplace platforms. Marketplace has really kicked off in recent years as a way for individuals to get rid of existing belongings. However, competition on these platforms have also skyrocketed, with buyers having an overwhelming number of goods to choose from. These platforms typically standardize fonts and limit color options. Therefore, sellers need to get creative and figure out different ways to stand out.

We decided to focus on text formatting and understanding whether or not that played a role in one's interest in the product. By influencing the formatting of the descriptive text, we hope to see a change in engagement and overall intent to purchase. Introducing bold or italicized text may better draw attention to specific words and capture the attention of the audience. Adding EMOJIS or symbols can also elicit emotions of the purchaser. As online sales continue to rise, we would like to identify ways to improve sales on an online

platform.

We do reckon that capitalizing all the words in an AD can also have a potential negative impact. Users may find it too blaring, pushy, or otherwise cheap, and may choose to ignore the AD.

Our aim is to study through an experiment the impact of the text formatting with the realization that treatment has as much chance to negatively impact as much as it can result in more clicks.

Research Question

Our primary research question is to understand if text formatting of an online car ad affects the amount of interest in the product itself. We honed in on the car resale market, because we wanted to target consumers who would actually read the ad description; since cars have steeper price points, the buyer will actually read the description in order to learn more about the mileage, make, model, feature, and condition of the car. Additionally, several details of the card can't be conveyed through pictures alone.

This experiment is also testing whether eliciting emotion may affect how a product sells. Fundamentally, humans are social, emotional creatures who yearn for connection. Experiments can be designed to test this mechanism, such as utilizing the description of a food item with either mundane (control) and exciting (treatment) adjectives in a service ad. There is a particular threshold for emotion, in that minimal amounts may not be discernible. Our experiment isn't testing a threshold, but rather a context. Will eliciting emotion specifically the description of a car ad drive greater marketability?

Hypothesis

When looking for a large purchase, such as a car from a private seller, a buyer generally wants to purchase a car from someone who is honest and trustworthy. The potential buyer may use formatting, may be subconsciously sometime, as a way of subtly detecting the character or type of seller the person is before engaging in private conversation for the potential sale. We hypothesize that if the details section of the ad is in all caps, it could give the impression of a pushy used car salesman and deter potential buyers.

Experiment Design

Experiment Overview

We conducted a clustered one-factor experiment. Our four members posted ten different ads on *Craigslist*: five control ads with standard capitalization of terms used to describe the car and five treatment ads in which these descriptive terms were all capitalized. We split the treatment and control among ten cities such that the chance of a person in one region seeing an AD other than the one we targeted is nearly zero. In addition, if we posted both control and treatment ad in each location one of the ADs would have been flagged as duplicate, which might render both ads as fraudulent. Thus, each location was exclusively either in treatment or in control.

As a potential outcome for comparison, we were interested in the number and rate of responses for each ad. Unfortunately, we were unable to track the number of visitors for each ad in order to discern response success rates. Instead, we will look at the number of responses received during the first three weeks of the post date for each ad. From these responses, we created an "interest-scoring" system that will provide further analysis for our regression. Each click generated an email to us which we then parsed to quantify the interest viewers expressed in the AD. This is described in some more in the section **Data Collection and Measurements**.

Each post will have the same picture, content, and information. See Figure 1 for a sample post.
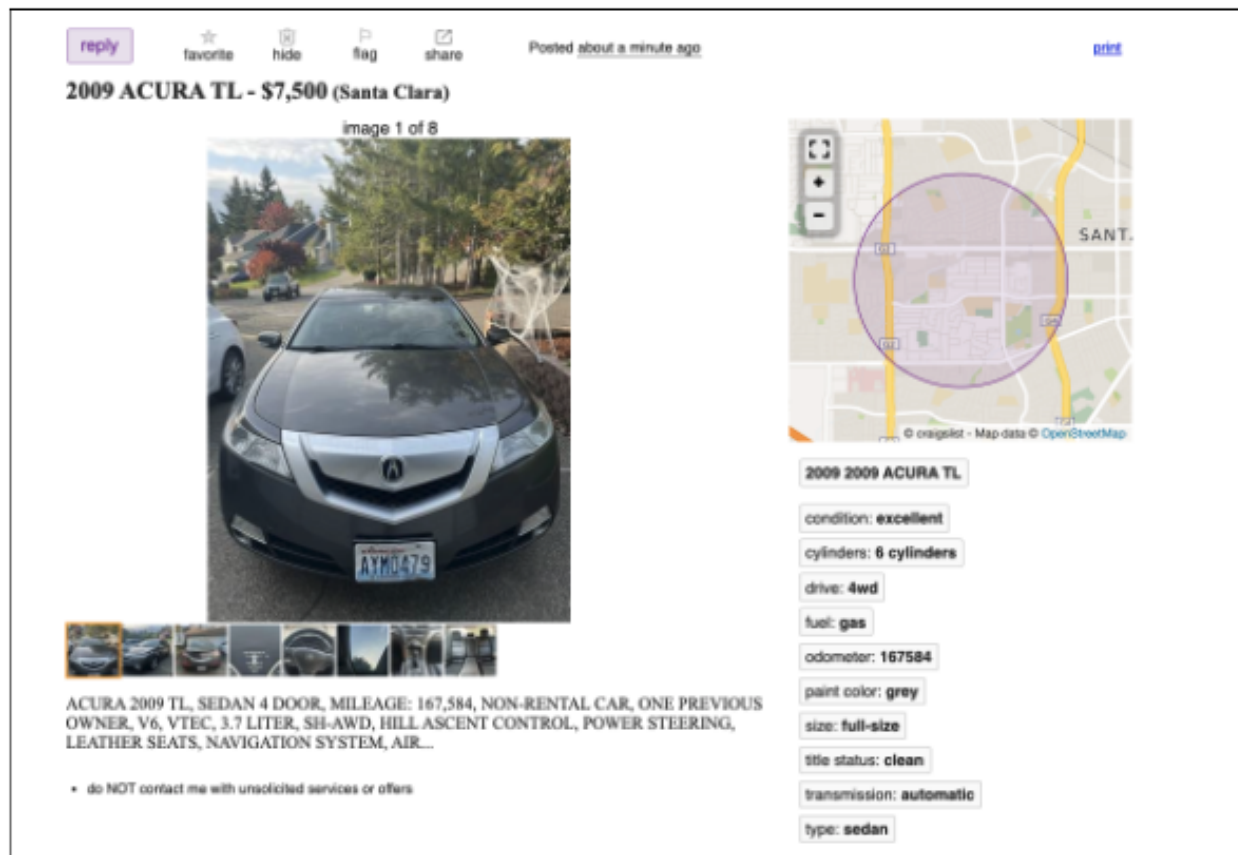
Figure 1: Treatment Post on Craigslist

Treatment and Control

Treatment and control both received identical information, content, and photos. However, what differed was the text formatting. The treatment received text in all uppercase capitalization but the control received standard English capitalization.

Each ad featured an Acura 2009 TL car with eight pictures encompassing different internal and external views. The description contains basic and necessary logistical and technical information about the vehicle. To determine the price for this used Acura 2009, we looked at other posts with the same model & make and similar mileage and took the average price of these posts. For the description, we tried to mimic example ads from Autotrader.

Group	Craigslist Description
Control	Acura 2009 TL, Sedan 4 Door, Mileage: 167,584, Non-Rental car, One Previous owner, V6, VTEC, 3.7 Liter, SH-AWD, Hill Ascent Control, Power Steering, Leather Seats, Navigation System, Air Conditioning, Heated Seats, Anti-Theft System, Great Condition with Minor Wear, Backup Camera, Moonroof
Treatment	ACURA 2009 TL, SEDAN 4 DOOR, MILEAGE: 167,584, NON-RENTAL CAR, ONE PREVIOUS OWNER, V6, VTEC, 3.7 LITER, SH-AWD, HILL ASCENT CONTROL, POWER STEERING, LEATHER SEATS, NAVIGATION SYSTEM, AIR CONDITIONING, HEATED SEATS, ANTI-THEFT SYSTEM, GREAT CONDITION WITH MINOR WEAR, BACKUP CAMERA, MOONROOF

Table 1: Control vs. Treatment Text Formatting

Data collection and Measurements

The data flow is captured in the picture below:
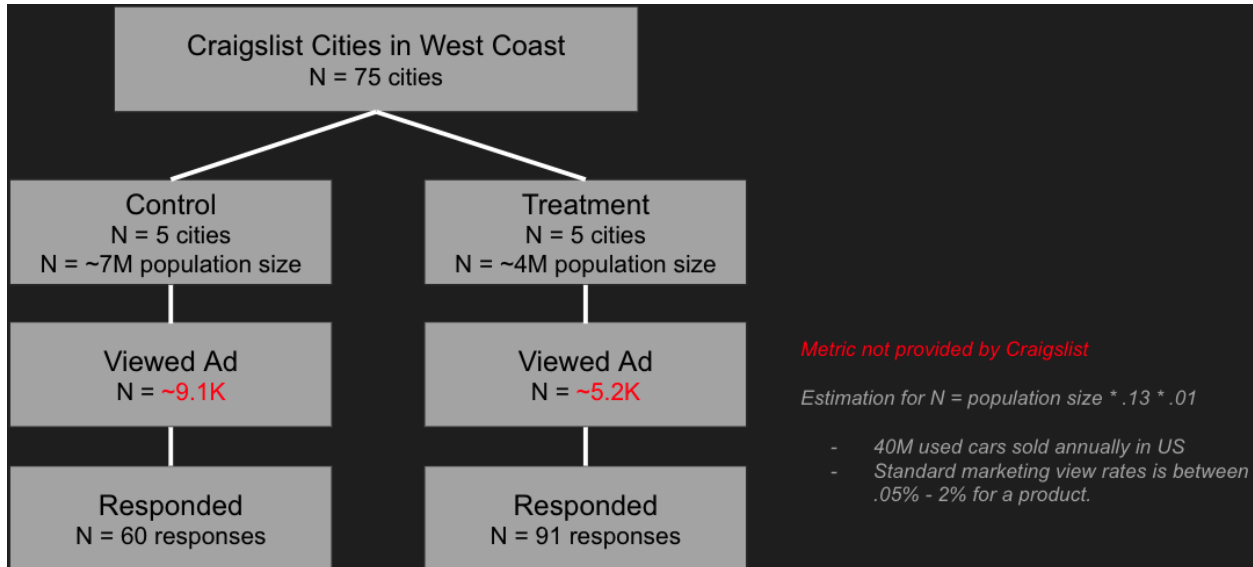
We posed the ad to *Craigslist* starting from Oct. 17th, 2021. Our measurement was done with email responses from those who clicked. We each had our private email listed as the contact email for each of our AD placement. When users clicked on the AD for a specific region we received an email, and we immediately started populating a row for each such response.

When email responses were received we transferred details such as the date and time when the response were received, the unique user name (assigned by *Craigslist*), the actual user-name, and we also parsed for

various interest levels. The information we gather from the email is explained in the table below:

Interest Criteria	Description
Contact Provided	Provided contact information - phone number, name, email
Ready to buy	Clearly express a desire to buy than just exploratory
Want now	Expresses a sense of urgency to buy the car
Ready to pay.	Makes an offer or indicates they have funds to buy
Set appointment	Offers to meet to examine the car and/or close the deal
High interest	Clicked the AD more than once; multiple queries

Table of Interest Criteria

Each of the "Interest Criteria" got a 1 if the email from the user supported that criteria. Otherwise the criteria got a score of 0. Some of the email scored on multiple criteria. For instance, one of the emails had the respondent's cell phone number, and the respondent wanted to set an appointment to see the car. We also created a composite "interest_final" column in the data that carried the sum of these individual criteria. The Google sheet was formatted to be able to compute these sums automatically. We also added a routine to convert the date and time of the email response to the corresponding "date-time" value which carried the number of seconds elapsed from Jan. 1st, 1970 mid-night.

The email below has interest_final value of 3. The person clicked, provided contact, and wanted an appointment.

This email has interest_final value of 5. The person clicked the AD, ready to pay (cash), set appointment, want to buy now, and provided contact.

In addition, we added the median household income, population, and distance from Seattle (where the car is registered) as additional columns. The idea was to use these additional parameters to see if the population density or proportional income gave additional insights.

Randomization

We identified ten metropolitan cities on the West Coast to post our ads in. We decided on a clustering randomization approach to avoid potential spillover. By posting both the treatment and control in the same city, we run the risk of having an individual see both ads and question its legitimacy. The ad likely could have been flagged by the *Craigslist* platform as duplicate posts within the same city. Therefore, our design ensured everyone within the same city was assigned to the same group.

One potential caveat is that different cities may have different utilization of *Craigslist*. Therefore, we decided to limit this experiment to the more populous cities on the West Coast that all have relatively high activity on *Craigslist*. We also selected cities that have relatively similar rates of driving. For example, we decided to post in Santa Clara instead of San Francisco, because those living in SF have access to good public transportation and high parking prices.

In retrospect, it would have been best to ensure that the total population for the treatment (4M) and control (7M) were relatively similar in size. It would have been a more rigorous randomization method for a pairing of treatment and control to be selected with a similar population. However, we will account for these population differences through normalization in our analysis.

Location	Treatment	Date of Posting	City Population
Seattle, WA	1	10/17/2021	724,305
Phoenix, AZ	0	10/18/2021	1,633,000
Los Angeles, CA	0	10/18/2021	3,967,000
Santa Clara, CA	1	10/18/2021	1,918,880
Reno, NV	1	10/24/2021	246,500
San Diego, CA	0	10/24/2021	1,410,000
Portland, OR	1	10/25/2021	645,291
Boise, ID	0	10/25/2021	226,115
Tucson, AZ	1	10/31/2021	541,482
Spokane, WA	0	10/31/2021	217,353

Table 2: Posting Locations

Observations and Outcomes

The data we collected breaks by location as indicated by the following figure.
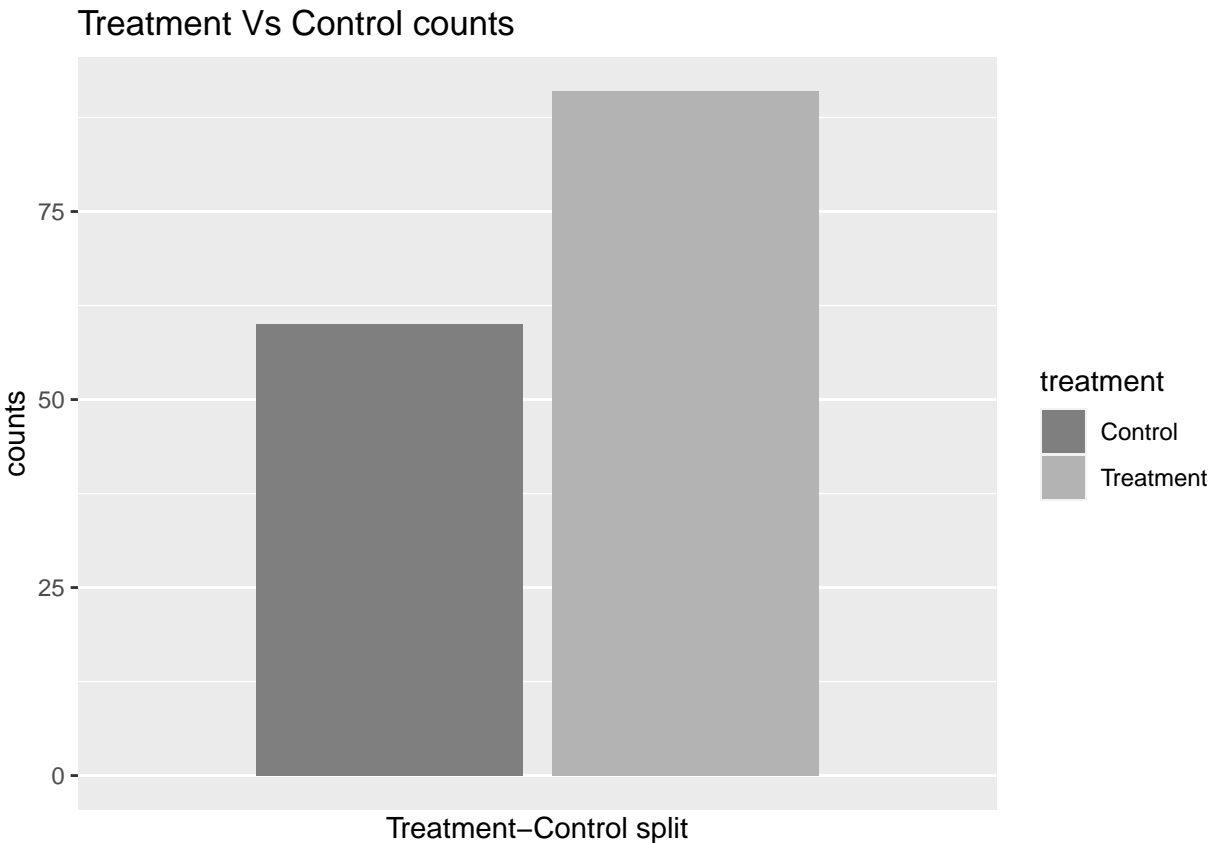

Response from each city
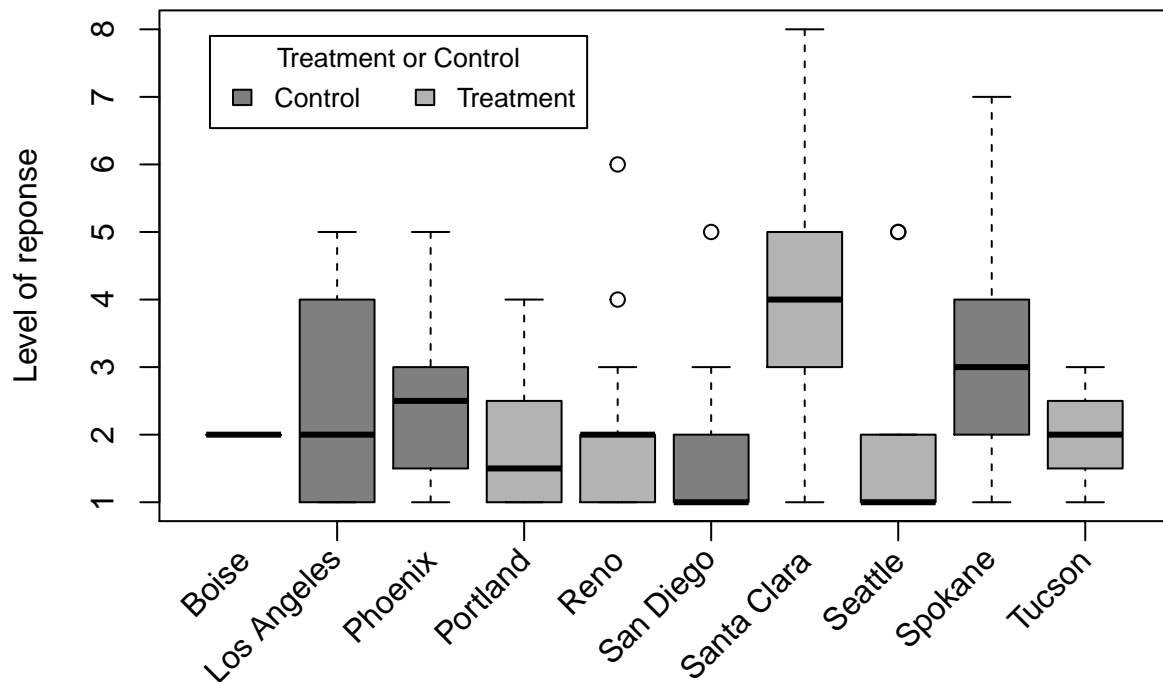
Note

that Boise is under the "Control" group. We don't see Boise's state clearly because there was only 1 response from Boise.

We also show the consolidated data broken by treatment and control.

Treatment Vs Control counts

Our outcome measure was the number of responses received for each *Craigslist* post and a point-based interest scale with the following criteria displayed in the table "*Table of Interest Criteria*", in the section **Data collection and Measurements**. We will sum the value of each criteria to calculate an integer point value greater than 0.

We observed several instances of multiple entries from buyers across almost all locations. For example, a user named "Robert Storseth" of Spokane replied four times from 10/31/2021 to 11/03/2021 with the response '6000 cash?'. To account for multiple entries, each email they sent will be calculated as a point for the number of messages sent. In this case, Robert received 4 points from this category. If the messages change for each subsequent email, we will only concern the criteria met in the very first message sent.

We also noticed that the ad response decayed with time quite precipitously. The 1st few days saw the most response, after which the responses went down to low single digits, and to zero or one soon after that. Older ads may likely either convey to a buyer that the information is outdated, in which an item has already been

sold, or it could be fake, in which an ad has been and/or is misleading. In addition, *Craigslist's* posting algorithm also biases towards new posts; they display ads in the order that it was posted from most recent to least.

We captured the response as a function of time as below:

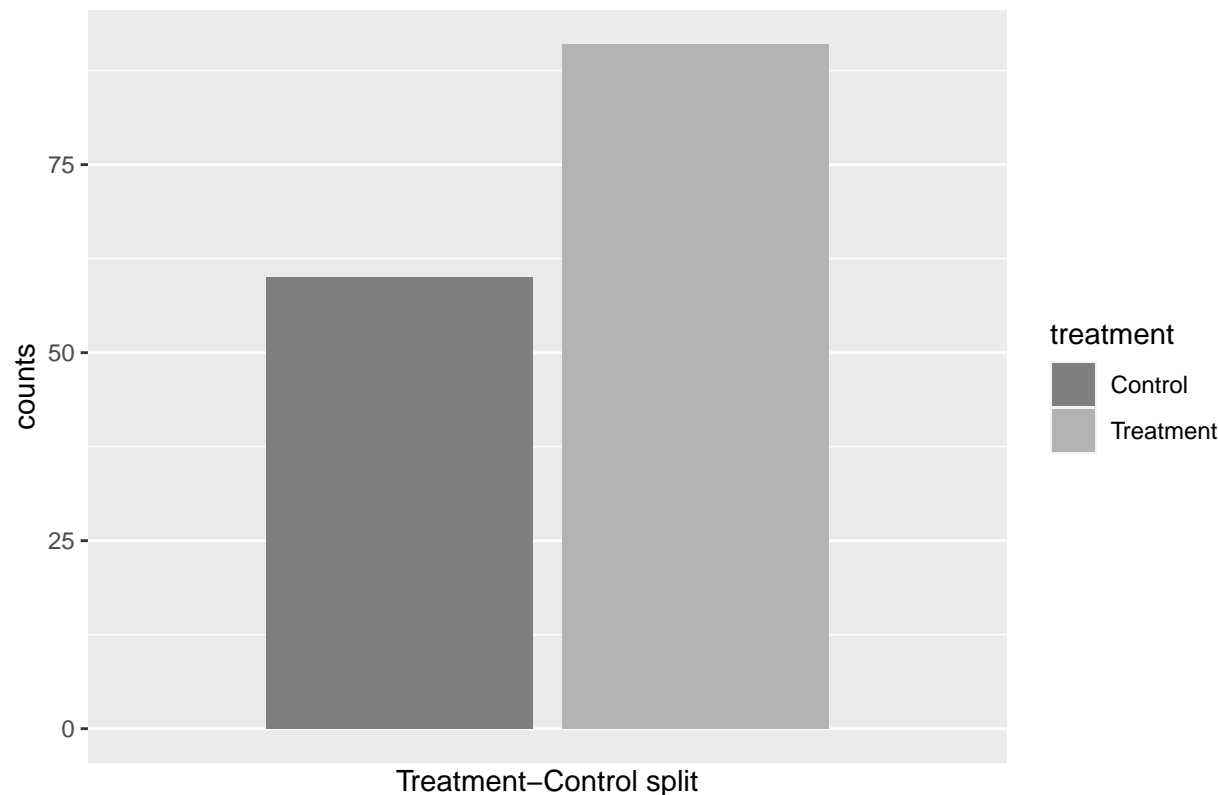## AD response as a function of date
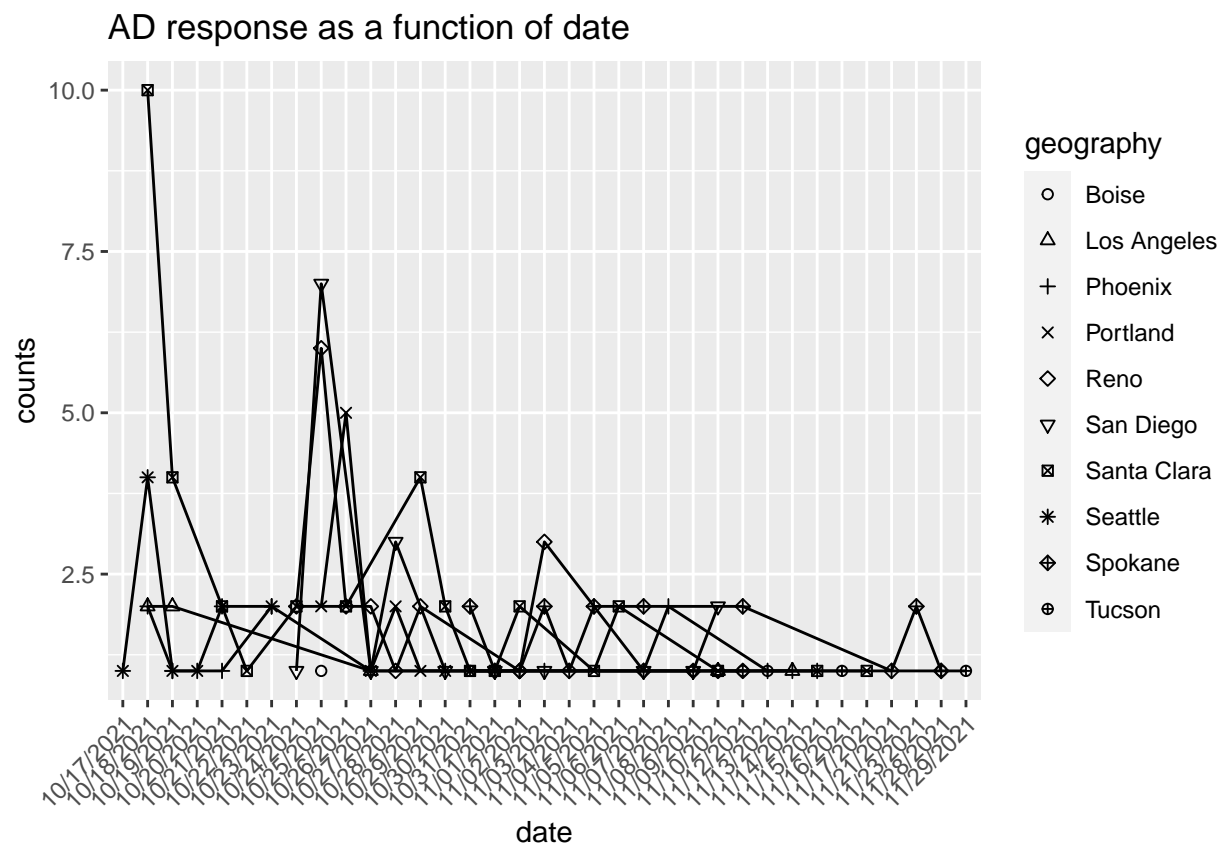
A potential issue that was uncovered during our unveiling of treatment and control was that several images of the car included a Washington state license plate. This could be a bias that potentially affected buyers' outlook of the sale as fraudulent such as those from cities outside Washington, such as Arizona or Nevada. This certainly may have affected the number of responses that these cities received. Our counteraction of this was choosing and alternating treatments for pairs of cities in close proximity to each other such as Tucson/Phoenix or Seattle/Spokane.

Results

With the analysis from 151 responses we gathered we wanted to answer if our hypothesis that all-capitalization impacts the response to a used-car online ad. While we just wanted to observed the response we were of the opinion that all-capitalization may turn users away from the ad, and that the

control ad (the ones with normal level of capitalization) will attract more statistically significant response. Numerically, we received more response from locations that were part of the treatment group, even though the population size of the control group was approximately 7.4M, while that of the treatment group was 4M.

Treatment Vs Control counts

This begged the question if treatment had any effect on the number of responses. A simple regression of the response count as a function of the treatment (see sub-section **Regression of response by treatment** under the **Regression** section) showed that the treatment had no statistically significant effect.

We then shifted attention to the "interest_final" variable. This variable captured the interest level of a person who clicked the ad. Based on the email response we assigned a 0 or 1 to various interest levels as explained in the sections **Data collection and Measurements** and in the section **Observations and Outcomes**.

We conducted a t-test of the "interest_final" split by treatment and control. The t-test came out to show that the Null Hypothesis holds good - i.e the interest_final score distribution isn't affected by the treatment assignment. The subsequent regression yielded a result consistent with the t-test. These are described under sub-section **Does interest level respond to treatment?** in the **Regressions** section.

We suspected if the vehicle being registered in Seattle, WA had any impact on the response in Seattle and in other cities. When online users buy a car from an unknown individual it's reasonable to suspect if the

local registration mattered or not. We found that in the type of resposne and in the regressions the license plate being that of Seattle, WA and the distance from Seattle didn't matter.

With added information of the population density and mean income for each location we conducted a series of regressions adding one covariate at a time. See sub-section **Add more covariates** in the **Regressions** section. We observed that in the model that regressed interest_final against "treatment", "population_density", "distance", and "mean income" we found that the income was the only variable that had a statistically significant effect. Every \$1000 increase in income contributes approximately 0.04 to the interest_final outcome.

We then plotted the mean interest_final level against mean_income level.
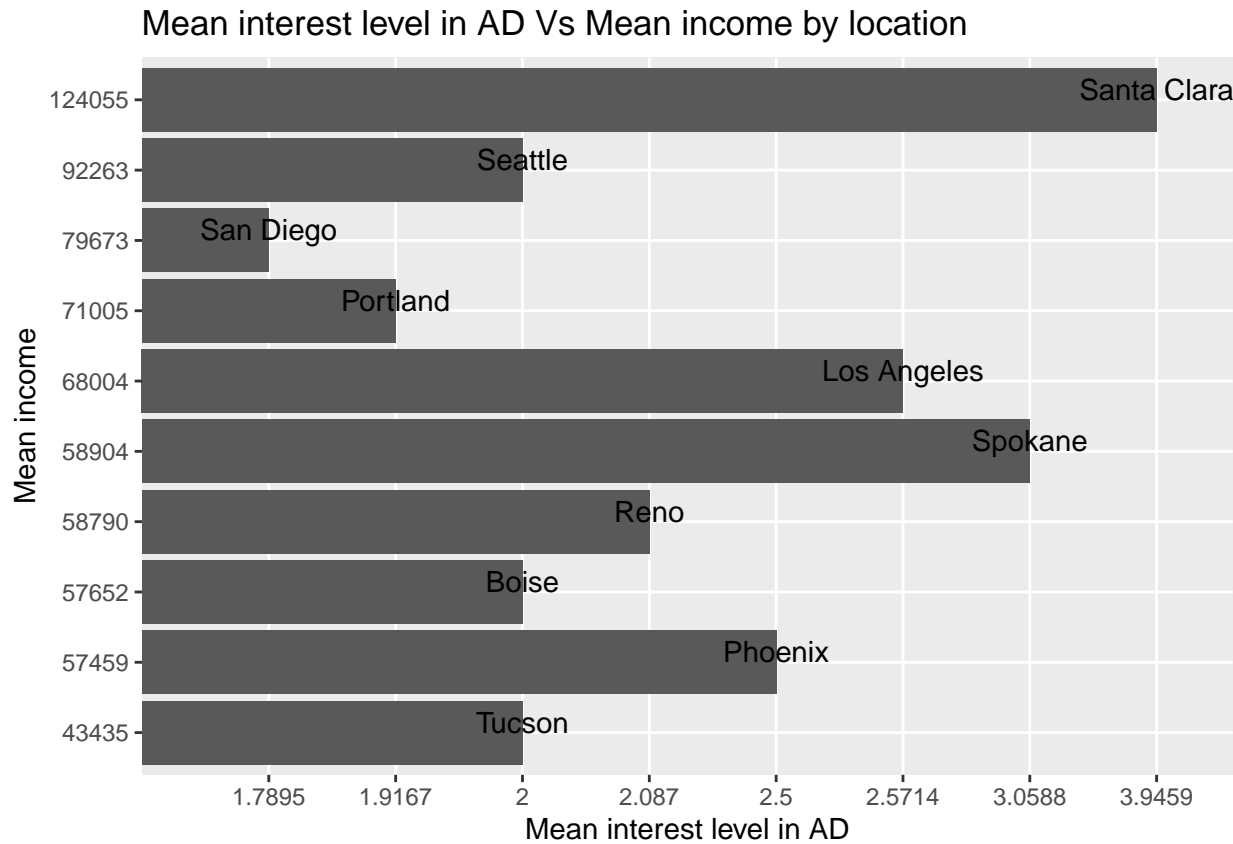
While the graph isn't monotonically showing higher interest level as a function of Mean Income it appears that the income has an overall impact on the response to the ad. Perhaps it's the combination of income and population density. However, we don't have results that substantiates such a conclusion.

Regressions

Regression of response by treatment

Let's first compare the number of responses from each of the locations (10 locations in total) based on the treatment or control the location was assigned to.

```
## 
## =========================================================
##                           Dependent variable:
##                      ----------------------------
##                                  counts
##                              (1)          (2)
## --------------------------------------------------------
## treatment                   6.200        6.200
##                            (7.101)      (6.351)
## 
## Constant                  12.000***    12.000***
##                            (3.841)      (4.491)
## 
## --------------------------------------------------------
## Observations                 10           10
## R2                         0.106        0.106
## Adjusted R2               -0.005       -0.005
## Residual Std. Error (df = 8)  10.042     10.042
## F Statistic (df = 1; 8)    0.953        0.953
## =========================================================
## Note:                      *p<0.1; **p<0.05; ***p<0.01
```

While the treatment shows a positive coefficient we don't see the effect of treatment as statistically significant. Neither Robust Standard Error (1) nor Cluster Standard Error (2) changes the outcome. Note that negative adjusted R2. This implies insignificance of the the explanatory variable. Likely a large sample size (many more locations) might help.

Does interest level respond to treatment?

The "interest_final" variable captures the sum of the scores of various interest levels. (See section **Data collection and Measurements** for explanation). We now examine if the variable "interest_final" records

any treatment effect using both t-test and a linear regression.

```
##
##  Welch Two Sample t-test
##
## data:  d[treatment == 1, interest_final] and d[treatment == 0, interest_final]
## t = 1.4442, df = 134.35, p-value = 0.151
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1362814  0.8740103
## sample estimates:
## mean of x mean of y
##  2.802198  2.433333

##
## ===============================================
##                          Dependent variable:
##                      ---------------------------
##                             interest_final
## ---------------------------------------------------
## treatment                      0.369
##                               (0.257)
##
## Constant                      2.433***
##                               (0.192)
##
## --------------------------------------------------
## Observations                    151
## R2                             0.013
## Adjusted R2                    0.007
## Residual Std. Error     1.566 (df = 149)
## F Statistic             2.007 (df = 1; 149)
## ===============================================
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

Clearly, both the t-test and the linear regression show no treatment effect that is statistically significant.

Add more covariates

```
## 
## ===============================================================================
##                                        Dependent variable:
## 
##                      ----------------------------------------------------------
##                                            interest_final
##                            (1)                  (2)                  (3)
## -------------------------------------------------------------------------------
## treatment                 0.378                0.170               -0.269
##                          (0.289)              (0.278)              (0.282)
## 
## population_density                           0.0002***            -0.0002
##                                              (0.0001)             (0.0001)
## 
## distance                 0.00003               0.0002              -0.0002
##                          (0.0003)             (0.0003)             (0.0003)
## 
## income                                                            0.00004***
##                                                                   (0.00001)
## 
## Constant                 2.407***             1.504***             0.995*
##                          (0.389)              (0.531)              (0.508)
## 
## -------------------------------------------------------------------------------
## Observations              151                  151                  151
## R2                        0.013                0.068                0.170
## Adjusted R2               0.00001              0.048                0.147
## Residual Std. Error  1.571 (df = 148)     1.532 (df = 147)      1.450 (df = 146)
## F Statistic          1.001 (df = 2; 148) 3.548** (df = 3; 147) 7.478*** (df = 4; 146)
## ===============================================================================
## Note:                                           *p<0.1; **p<0.05; ***p<0.01
```

The output above shows that in all cases treatment effect is statistically insignificant. Distance from Seattle (where the car is registered) also didn't matter. Model (1) carried all its weights in the Constant. Model (2) pushed part of weight from constant into population density. Model(2) claims that every additional 100 people per square mile add 0.02 to the final interest outcome. Model (3) shows that the only variable that matters is the median income of the location. Every $1000 increase in income contributes approximately 0.04 to the interest_final outcome.

Conclusions

The study set out to find the effect of capitalizing wordings on an online ad for a used-car. Our assumption, when we started, was that all-capitalization will impact clicks on our ads, likely negatively. Our experiment didn't find any causality between the word formatting and the response level to the ad.

As we considered additional factors that might influence people desire to buy a used car we found that mean income may have a role to play. Other factors such as where the vehicle is registered and the population density of the location had no statistically significant role to play, in the final analysis.

As described in the next section a more robust A/B test framework may be the right framework to study the impact of variation in ads.

Limitations and Future Enhancements

One of the biggest limitations to this experiment is that *Craigslist* does not provide additional metrics on the number of views and reach per post. Therefore, we are basing this entire experiment on the number of responses received. If we had additional information, we would be able to calculate a more accurate response rate for the treatment and control. In addition, we do not have any demographic information about the individuals who responded to our ad.

We also wanted to see the effects of text formatting on different online marketplaces. However, we faced several platform limitations for some of these sites. One option was to create a post on Facebook Marketplace and then turn the post into an ad by boosting it when publishing. We will create two posts (one treatment and one control). We are concerned that FB will flag duplicate posts, which is why we are only limiting it to two posts by two accounts in separate locations. Unfortunately, we are not able to use FB's built-in A/B testing functionality because it requires a business account, company registration, and a company page to do so. We are concerned that having a bare-bones page (with no photos or history of posts) would be

suspicious and deter users from reaching out about the post. In addition, we did not want the personal information of the fake Facebook profile to detract from the content of the ad. Therefore, we opted to forgo this option. *Autotrader* was another marketplace we originally considered. However, the site has a steep fee for each post and requires proper documentation, like VIN numbers and license plate number, for each car in order to authenticate the ad. Given that we were not actually planning on selling multiple used cars, we were unable to pull this information together. In addition, this would limit the number of ads we were able to pull. Ultimately, we choose to use *Craigslist*, because posting on the platform itself introduces the least amount of confounding variables. *Craigslist* also does not flag duplicate posts with the same content and photos, which allows us to post in more cities around the same time without any issues. Given more time, we would ideally post on more platforms.

We would also like to call out the current used-car shortage happening within the United States. If we had the data, we could have done a deeper dive analysis into understanding the cities with a relatively similar shortage rate and target those locations.

One interesting future enhancement that we did not have the opportunity to conduct is advertising cars of varying price points and brands. We could have even simplified this experiment and did a 2x2 factorial design with text formatting (normal caps or all caps) and price points (high or low). However, we did not have used cars or photos and consequently, could not have published fake ads for. However, if we did have time to collect this information, we could have offered a $100 gift card to those willing to share pictures of their vehicle to us and be part of this study.