**Anmol More – 11915043**

**Please Note :**

All analysis has been done in single excel file with multiple sheets

Additional graph plotting and analysis done in python

1. **Is Online Advertising effective for Star Digital ?**

**No, Online advertising under given conditions, where advertiser cannot control between sites 1-5, is not effective for star digital. I have created three models in support of this hypothesis.**

We can see from **Model 1** in excel that odds of purchase increases by 0.019 when exposed to online ad of star digital when compared to customers who were shown charity ads.

However, our confidence in above statement is low, as significance level (p value) is high at 0.06, which means that there is 0.06 probability that showing star digital ad is not effective for star digital, so we go on doing further analysis.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | *Coefficients* | *P-value* |
| Intercept | 0.48569277 | 0 |
| test | 0.01918646 | 0.06135393 |

**On further tests using Model 2 –**

1. Since we have two set of ads, I created two sets of users, those who viewed ads for either charity or star digital on website 1-5 (Shown\_1-5), and those who were shown ads on website 6 (Shown\_6)
2. Also, I created two interaction variables test\*shown\_1-5 and test\*shown\_6

On running regression as seen in Model 2 effect on multicollinearity has been verified using correlation matrix –

Here are observations **–**

1. We find that, test set has p value as 0.728, so it didn’t matter whether user saw a star digital ad or charity ad, purchase decision was not affected by viewing star digital ad.
2. When user saw a star digital ad on any of the websites in 1 to 5, it did not increase his/her chances of purchase, as interaction variable sum coefficient comes to almost 0 (-0.0035237+0.00297941)
3. Above statement holds equally true for website 6 – Viewing a star digital ad on website 6 didn’t improve the chances of purchase, as coefficient sum is close to zero (-0.0035237 + 0.00358612)

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | 0.39284574 | 2.674E-182 |
| test | -0.0035237 | 0.72814854 |
| Shown\_1-5 | 0.141952 | 3.8237E-56 |
| Shown\_6 | -0.0032041 | 0.70887601 |
| test\*shown\_1-5 | 0.00297941 | 1.6933E-66 |
| test\*shown\_6 | 0.00358612 | 7.0408E-12 |

**Did further analysis using model 3 –**

Created new variables – SUM\_imp\_1-5 (Sum of all impressions across website 1 to 5) and created interaction variables with test set

Observations –

Increasing no of impressions for star digital ad didn’t made a significant difference in purchase decision as p value for either test test\*Sum\_1-5, test\*imp\_6 is > even 0.1 at 0.25, 0.14 and 0.159 respectively

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | 0.4653462 | 0 |
| test | 0.01211166 | 0.25894704 |
| imp\_6 | 0.00089974 | 0.32634767 |
| SUM\_1-5 | 0.00307801 | 9.1087E-11 |
| test\*SUM\_imp\_1-5 | 0.00073011 | 0.14724555 |
| test\*imp\_6 | 0.0014738 | 0.15999638 |

1. **Is there a frequency effect of advertising on purchase ?**

**Yes, there is a frequency effect of advertising on few websites but not all**. I have created three models in support of this hypothesis.

Let us first check a simplistic model - **Model 3** –

As shown in question 1 above as well, just increasing the number of impressions for star digital ad under two ad groups, didn’t made a significant difference in purchase decision as p value for test set flag and interaction variables both are quite high > 0.01

However, we go on further analyzing effect of no of impressions on each individual website on purchase through **Model 4 –**

For each website, I created interaction variable with test set and ran a regression. Observations –

1. We can see that for website 1 and 3 p value is not significant hence we can say frequency effect is not valid for website 1 and 3
2. For website 2, 3 and 6 for each additional impression, probability of purchase increases by 0.002, 0.0131 and 0.002 ie. chances of purchase increases by 0.2%, 1.3% and .2% times respectively.

More no of impressions on website 5 has negative effect and brings down chances of purchase by 4.4%.

1. Since we don’t have option to choose between website 1-5, we can say only website 6 has positive impact on purchase with increasing no of impressions

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | 0.48569277 | 0 |
| test | -0.0105746 | 0.29743541 |
| test\*imp\_1 | -0.0011851 | 0.06478415 |
| test\*imp\_2 | 0.00228475 | 2.716E-20 |
| test\*imp\_3 | 0.02288016 | 0.43532325 |
| test\*imp\_4 | 0.013151 | 3.624E-128 |
| test\*imp\_5 | -0.0448589 | 1.0882E-15 |
| test\*imp\_6 | 0.00224713 | 9.8274E-06 |

**On analysis on data, we see that all impression variables are right skewed.**

Percentile charts along with box plots and histograms can be seen in python code.

Created a new variable with natural log of each of the impression variables to normalize the values of impression variables

| Percentiles | **imp\_1** | **imp\_2** | **imp\_3** | **imp\_4** | **imp\_5** | **imp\_6** |
| --- | --- | --- | --- | --- | --- | --- |
| 0.100 | 0.000 | 0.000 | 0.000 | 0.0 | 0.0 | 0.000 |
| 0.200 | 0.000 | 0.000 | 0.000 | 0.0 | 0.0 | 0.000 |
| 0.800 | 0.000 | 3.000 | 0.000 | 0.0 | 0.0 | 2.000 |
| 0.900 | 1.000 | 6.000 | 0.000 | 4.0 | 0.0 | 4.000 |
| 0.950 | 4.000 | 14.000 | 0.000 | 9.0 | 0.0 | 6.000 |
| 0.990 | 18.000 | 57.000 | 2.000 | 30.0 | 1.0 | 20.000 |
| 0.999 | 80.792 | 182.792 | 16.396 | 83.0 | 7.0 | 83.698 |
| Max | 296 | 373 | 148 | 225 | 51 | 404 |

I created two more models without interaction variables in **model 5 and 6**, Observations –

1. We can see again that, more no of impressions on website 2, 4 and 6 only are significant and cause positive impact on purchase
2. Here again, more no of impressions on website 5 result in negative effect on purchase

**Model 5 coefficients and p values**

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | 0.42735862 | 0 |
| test | 0.00743817 | 0.46205861 |
| ln(imp\_1) | -0.0082183 | 0.1901497 |
| ln(imp\_2) | 0.0281314 | 1.2219E-16 |
| ln(imp\_3) | -0.0366029 | 0.04249433 |
| ln(imp\_4) | 0.18513328 | 0 |
| ln(imp\_5) | -0.2147822 | 2.1275E-19 |
| ln(imp\_6) | 0.02681024 | 1.093E-09 |

**Mode 6 coefficients and p values**

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | 0.46205292 | 0 |
| imp\_1 | -0.0006036 | 0.31256102 |
| imp\_2 | 0.00228728 | 2.5101E-22 |
| imp\_3 | -0.0050933 | 0.01509535 |
| imp\_4 | 0.01265974 | 8.447E-137 |
| imp\_5 | -0.0439269 | 8.0165E-16 |
| imp\_6 | 0.00190906 | 1.6664E-05 |
| test | 0.01412273 | 0.16609698 |

1. **Where should they advertise?**

**If given a choice they should advertise only on website 4. For every additional impression on website 4, chances of purchase increases by 0.2%. Explanation below –**

**Also, to have positive impact on purchase, they shouldn’t be advertising on other websites, as they ultimately result in negative or zero affect on purchase.**

Where not to advertise –

star digital shouldn’t be advertising under two group of websites, where they have only two choices either show ads in 1-5 or on website 6.

We have seen in **model 2**, those who were in test group don’t show any significant difference from purchase of those who were in control group. Since, p value for both test variables is > 0.1 ie 0.72. Also, combined interaction effect on purchase effect comes close to zero (-0.0035237+0.00297941) for websites 1-5 and for website 6 (-0.0035237 + 0.00358612)

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | 0.39284574 | 2.674E-182 |
| test | -0.0035237 | 0.72814854 |
| Shown\_1-5 | 0.141952 | 3.8237E-56 |
| Shown\_6 | -0.0032041 | 0.70887601 |
| test\*shown\_1-5 | 0.00297941 | 1.6933E-66 |
| test\*shown\_6 | 0.00358612 | 7.0408E-12 |

**Again in model 3 we can see that**, combined effect of impressions across website 1-5 or impressions through website 6 on test set group is not significant as test variable and interaction variable all have p values greater than 0.1

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | 0.4653462 | 0 |
| test | 0.01211166 | 0.25894704 |
| imp\_6 | 0.00089974 | 0.32634767 |
| SUM\_1-5 | 0.00307801 | 9.1087E-11 |
| test\*SUM\_imp\_1-5 | 0.00073011 | 0.14724555 |
| test\*imp\_6 | 0.0014738 | 0.15999638 |

Now, let is see **model 4, for** Combined effect of being on test set and impressions on –

1. Website 2 – -0.0105746+0.00228475 = -0.008
2. Website 4 – 0.013151-0.0105746 = 0.002
3. Website 6 – 0.00224713-0.0105746 = -0.008

Advertisements on websites 1 and 3 doesn’t had any significant affect on purchase. Also website 5 has negative impact on purchase. Now seeing the combined effect on other websites, we can say that, only advertisements on website 4 is having positive impact

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | 0.48569277 | 0 |
| test | -0.0105746 | 0.29743541 |
| test\*imp\_1 | -0.0011851 | 0.06478415 |
| test\*imp\_2 | 0.00228475 | 2.716E-20 |
| test\*imp\_3 | 0.02288016 | 0.43532325 |
| test\*imp\_4 | 0.013151 | 3.624E-128 |
| test\*imp\_5 | -0.0448589 | 1.0882E-15 |
| test\*imp\_6 | 0.00224713 | 9.8274E-06 |

1. **Why experiments? Are there other ways to achieve this ?**

Experiments give power to measure the causal effect of display advertising on sales conversion. Instead of just relying on click through rates or impressions, by creating random group of users and measuring the increase in conversions due to seeing the advertisements compared to not seeing the advertisements can help to determine the value of advertisement.

Identifying this incremental impact on sales, without treatment and control group is not possible.

Other ways to achieve this could be –

1. Advertising through influencers – Other than experiments, star digital can entertain influencers who can talk about their product and will ultimately result in more no of impressions through people reading out their vblogs, blogs, articles and result in purchase decisions
2. Social media analysis – We can improve the social media presence of star digital and have them more responsive to consumer comments and complaints, which might result in increase brand perception and result in more no of customer acquisitions
3. Referral purchases – Give incentives to existing customers for referrals, which can increase sales
4. **What is the criterion for setting this up?**

Important criterial to setup up experiments are –

1. Randomly assigning users in control group, to which advertisements will not be served and test group to which advertisements will be served
2. Identifying opportunity cost of not serving the advertisement to users, who could have resulted in purchase by showing the advertisement
3. Identifying the fraction of users in control and test group based on baseline conversion rate, campaign reach, minimum lift and power of experiment
4. Ideally making sure that only difference between control and test group is treatment, in this case showing the advertisements.