**Anmol More – 11915043**

**Please Note :**

* **Analysis has been done with regression models in excel (However regression models are statistically incorrect for binary target)**
* **Binomial generalized linear models and classification models are used below to prove hypothesis and arrive on solution based on python**
* **Additional graph plotting and analysis done in python**

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

**No, Online advertising under given conditions, is not effective for star digital. Models below show that putting a customer under test set alone doesn’t improves his chances of purchase. I have created five models in support of this hypothesis.**

We can see from **Reg Model 1 (Linear Regression, excel)** 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 |

**Since, we have binary target variables, regression is flawed technique here and instead logistic regression and classification models should be built for our hypothesis.**

**On further tests using Model 2 (Binomial GLM, python) –**

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | -0.0572 | 0.140 |
| test | 0.0768 | 0.061 |

p value is high at 0.06, which means that test set is not significant at significance value of 0.05. Hence we can say that effect showing star digital ad of star digital ads compared to customers who were shown charity ads on purchase decisions is insignificant.

**On further tests using Model 3 (Binomial GLM, python) –**

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

|  |  |  |
| --- | --- | --- |
|  | *Coefficients* | *P-value* |
| Intercept | -0.4732 | 0.000 |
| test | 0.0890 | 0.502 |
| Shown\_1\_5 | 0.6112 | 0.000 |
| Shown\_6 | 0.0120 | 0.909 |
| test\_Shown\_1\_5 | 0.1619 | 0.170 |
| test\_shown\_6 | 0.1258 | 0.256 |

Before running binomial GLM as seen in Model 3, effect on multicollinearity has been verified using correlation matrix –

Here are observations **–**

1. We find that, test set has p value as 0.502, 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 test\_Shown\_1\_5 p value > 0.05, hence not significant.
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 p value > 0.05.

**Did further analysis using model 4 (Decision Tree Classifier) –**

Below classification model using decision tree classifier clearly shows that, whether a user was in test group or not, It didn’t affected the purchase decision.

Instead purchase is dependent on whether they have seen a ad on website 6

**A close up of a logo

Description automatically generated**

**Further prove my analysis using Model 5 (Decision Tree Classifier) –**

Created new variables – SUM\_imp\_1-5 (Sum of all impressions across website 1 to 5) below decision tree clearly shows that purchase decision is not affected by being in test set

A picture containing map, text

Description automatically generated

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

**Yes, there is a frequency effect of advertising on few websites but not all. It looks like websites itself have an effect on purchasing rather than advertising decision**. I have created three models in support of this hypothesis.

Let us first check a simplistic model - **Model 3 and Model 5** (Sum of total impressions) above –

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 6 –**

============================================================================

coef std err z P>|z| [0.025

-----------------------------------------------------------------------------

Intercept -0.1927 0.043 -4.505 0.000 -0.277

test -0.0144 0.046 -0.317 0.751 -0.104

imp\_1 0.0171 0.012 1.426 0.154 -0.006

imp\_2 0.0120 0.005 2.554 0.011 0.003

imp\_3 -0.0328 0.016 -2.062 0.039 -0.064

imp\_4 0.1160 0.017 6.912 0.000 0.083

imp\_5 -0.1602 0.124 -1.292 0.196 -0.403

imp\_6 0.0035 0.004 0.827 0.408 -0.005

test\_imp\_1 -0.0288 0.013 -2.282 0.022 -0.054

test\_imp\_2 0.0032 0.005 0.635 0.525 -0.007

test\_imp\_3 0.2557 0.153 1.676 0.094 -0.043

test\_imp\_4 0.0751 0.018 4.069 0.000 0.039

test\_imp\_5 -0.3291 0.135 -2.446 0.014 -0.593

test\_imp\_6 0.0165 0.005 3.085 0.002 0.006

=============================================================================

**Model 7 (Decision Tree Classifier) –**

**A close up of text on a white background

Description automatically generated**

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

1. We can see that for website 3 and 5 there is a negative effect on advertising, Also due to advertising on website 5, it resulted in further negative effect.
2. For website 4 and 6 for each additional impression, probability of purchase increases by 0.0751 and 0.0165 ie. odds of purchase increases by 7.51 and 1.65 times respectively.

More no of impressions on website 5 has negative effect and brings down odds of purchase by 32%

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

**On analysis on data in excel, 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 |

1. **Where should they advertise?**

**If given a choice they should advertise only on website 4 and website 6. For every additional impression of star digital ad (ie. For customers in test set) on website 4, odds of purchase increases by 7.5%. And for every additional impression on website 6, odds of purchase increases by 1.65%. 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 3 and 6 above**, 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 0.5 and 0.7.
* Going by **model 6**, if we don’t consider significance, combined interaction effect of all websites 1-5 is -0.0288+0.0032+0.2557+0.0751-0.3291 = -0.02, Hence advertising in group of websites 1-5 without getting to choose one of them is hurting the brand with odds of purchase going down due to advertising.
* Going by Model 6 and Decision Tree Classifier in Model 7, effect of advertising on –
  + Website 1 – For every additional impression of star digital ad, odds of purchase went down by 2.8%
  + Website 2 and 3 – Advertising had no significant effect
  + Website 5 – For every additional impression of star digital ad, odds of purchase went down by 32%
  + Website 4 and Website 6 – Results are statistically significant and every additional ad impression resulted in 7.51 % and 1.65% increase in odds of purchase.

coef std err z P>|z| [0.025

-----------------------------------------------------------------------------

Intercept -0.1927 0.043 -4.505 0.000 -0.277

test -0.0144 0.046 -0.317 0.751 -0.104

imp\_1 0.0171 0.012 1.426 0.154 -0.006

imp\_2 0.0120 0.005 2.554 0.011 0.003

imp\_3 -0.0328 0.016 -2.062 0.039 -0.064

imp\_4 0.1160 0.017 6.912 0.000 0.083

imp\_5 -0.1602 0.124 -1.292 0.196 -0.403

imp\_6 0.0035 0.004 0.827 0.408 -0.005

test\_imp\_1 -0.0288 0.013 -2.282 0.022 -0.054

test\_imp\_2 0.0032 0.005 0.635 0.525 -0.007

test\_imp\_3 0.2557 0.153 1.676 0.094 -0.043

test\_imp\_4 0.0751 0.018 4.069 0.000 0.039

test\_imp\_5 -0.3291 0.135 -2.446 0.014 -0.593

test\_imp\_6 0.0165 0.005 3.085 0.002 0.006

=============================================================================

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 without advertising analysis 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.