Evaluating Cost-Effective Ad Placement for Star Digital

Executive Summary

This report evaluates the effectiveness of **Star Digital's online display advertising**. The primary objectives are to: 1. **Assess whether online advertising increases purchases**. 2. **Determine if ad frequency impacts conversion rates**. 3. **Identify which sites Star Digital should prioritize for advertising**.

Our analysis uses experimental data, statistical tests, and regression models to provide **data-driven recommendations** for optimizing advertising spend.

Data Preparation

```
data <- read_excel("M347SS-XLS-ENG.xls")
data = data %>% mutate(total_imp = imp_1 +imp_2 +imp_3 +imp_4 +imp_5 +imp_6)
```

Is Online Advertising Effective?

Checking Balance Between Control and Treatment Groups:

The experiment was **randomized by design**, meaning users were assigned to either the test or control group at the moment of ad serving. This should theoretically ensure that both groups are **statistically similar**.

To validate this, we conducted a **t-test on total impressions*:

```
t.test(total_imp ~ test, data = data)
```

```
##
##  Welch Two Sample t-test
##
## data:  total_imp by test
## t = 0.12734, df = 3204.4, p-value = 0.8987
## alternative hypothesis: true difference in means between group 0 and group 1 is not e
## 95 percent confidence interval:
```

```
##  -0.8658621  0.9861407
## sample estimates:
## mean in group 0 mean in group 1
##        7.929217        7.869078
```

Findings: The p-value (0.8987) indicates no significant difference in impressions between the test and control groups. This confirms that randomization was successful.

To ensure that our experiment has enough observations to detect a statistically significant effect, we conduct a **power test**:

```r
n_control <- sum(data$test == 0)
n_treatment <- sum(data$test == 1)

power_result <- power.t.test(delta = 0.2*0.5,
                             sd = 1,
                             sig.level = 0.05,
                             power = 0.8,
                             type = "two.sample",
                             alternative = "two.sided")

power_result
```

```
##
##      Two-sample t test power calculation
##
##               n = 1570.737
##           delta = 0.1
##              sd = 1
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
n_control
```

```
## [1] 2656
```

```r
n_treatment
```

```
## [1] 22647
```

- The computed sample size requirement was 1,570 per group, far below our actual sample sizes (2,656 in control and 22,647 in treatment).
- This indicates that our study is overpowered, meaning that if a real effect of advertising exists, we have sufficient data to detect it.
- Since our sample size greatly exceeds the required threshold, the results from our logistic regression models are statistically reliable.

Potential Assumptions and Limitations

While the experiment follows a standard A/B test framework, one potential concern is the **Stable Unit Treatment Value Assumption (SUTVA)**. This assumption requires that a user's purchase decision is influenced only by their own ad exposure, without spillover effects from other users.

In this experiment, potential violations of SUTVA include: - **Cross-device behavior:** If users see ads on one device but convert on another where tracking does not apply. - **Spillover effects:** If test group users influence control group users via word-of-mouth or organic brand awareness. - **Heterogeneous treatment effects:** Some users might be more influenced by ads based on prior exposure to similar brands.

While these effects are difficult to quantify, they should be kept in mind when interpreting the results. Future experiments could mitigate these risks by tracking cross-device behavior or segmenting users based on prior ad exposure.

Effect of Advertising on Purchases

A logistic regression is used to check whether being in the test group (advertised) increases the probability of purchase:

```
summary(glm(purchase ~ test, data = data, family = binomial))
```

```
##
## Call:
## glm(formula = purchase ~ test, family = binomial, data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.05724    0.03882  -1.474   0.1404
## test         0.07676    0.04104   1.871   0.0614 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 35073  on 25301  degrees of freedom
## AIC: 35077
##
## Number of Fisher Scoring iterations: 3
```

Results:

- The coefficient for test (0.07676, p = 0.0614) suggests a positive but marginally insignificant effect of advertising on purchases.
- This means advertising has some effect, but not strongly statistically significant.

Interaction with Ad Frequency

We check if ad exposure (total impressions) influences purchase probability:

```r
summary(glm(purchase ~ test * total_imp, data = data, family = binomial))
```

```
##
## Call:
## glm(formula = purchase ~ test * total_imp, family = binomial,
##     data = data)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.169577   0.042895  -3.953 7.71e-05 ***
## test           -0.013903   0.045613  -0.305    0.761
## total_imp       0.015889   0.002876   5.524 3.32e-08 ***
## test:total_imp  0.015466   0.003207   4.823 1.42e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 34190  on 25299  degrees of freedom
## AIC: 34198
##
## Number of Fisher Scoring iterations: 5
```

Findings:

- The coefficient for total_imp is positive and highly significant ($p < 3.32e\text{-}08$), confirming that more ad impressions increase the probability of purchase.

- The interaction term (test:total_imp) is also significant ($p < 1.42e-06$), suggesting that the effect of being in the test (advertised) group strengthens as ad impressions increase.

Does Ad Frequency Affect Purchases?

A logistic regression tests whether more impressions lead to more conversions:

```r
summary(glm(purchase ~ total_imp, data = data, family = binomial))
```

```
##
## Call:
## glm(formula = purchase ~ total_imp, family = binomial, data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.179392   0.014583  -12.30   <2e-16 ***
## total_imp    0.029201   0.001294   22.56   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 34212  on 25301  degrees of freedom
## AIC: 34216
##
## Number of Fisher Scoring iterations: 5
```

Findings:

- The coefficient for total_imp (0.029201, $p < 2e-16$) confirms ad frequency positively impacts purchases.

Diminishing Returns Analysis

We include a quadratic term to check if excessive ads reduce effectiveness:

```r
logit_quad <- glm(purchase ~ total_imp + I(total_imp^2), data = data, family = binomial)
summary(logit_quad)
```

```
##
## Call:
## glm(formula = purchase ~ total_imp + I(total_imp^2), family = binomial,
##     data = data)
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.406e-01  1.508e-02  -15.95   <2e-16 ***
## total_imp       4.369e-02  1.646e-03   26.54   <2e-16 ***
## I(total_imp^2) -1.158e-04  6.492e-06  -17.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 33959  on 25300  degrees of freedom
## AIC: 33965
##
## Number of Fisher Scoring iterations: 4
```

Findings:

- The negative quadratic term (-1.158e-04, $p < 2e-16$) confirms diminishing returns.
- There is an optimal number of impressions before ad effectiveness declines.

Which Sites Should Star Digital Advertise On?

Conversion Rate by Sites:

```r
site_conversion <- data %>%
  summarise(
    conv_1_5 = mean(purchase[(imp_1 + imp_2 + imp_3 + imp_4 + imp_5) > 0], na.rm = TRUE)
    conv_6 = mean(purchase[imp_6 > 0], na.rm = TRUE)
  )

print(site_conversion)
```

```
## # A tibble: 1 x 2
##   conv_1_5 conv_6
##      <dbl>  <dbl>
## 1    0.560  0.463
```

Findings:

- Sites 1-5 have a higher conversion rate (56%).
- Site 6 has a conversion rate of 46%.

How effective are Sites 1-5 compared to Site 6 ?

To compare whether advertising on Sites 1-5 collectively outperforms Site 6, we create an aggregated impression variable:

```r
data <- data %>%
  mutate(total_imp_1_5 = imp_1 + imp_2 + imp_3 + imp_4 + imp_5)

# Logistic regression for Sites 1-5 vs. Site 6
logit_sites <- glm(purchase ~ total_imp_1_5 + imp_6, data = data, family = binomial)

# Summary of the model
summary(logit_sites)
```

```
##
## Call:
## glm(formula = purchase ~ total_imp_1_5 + imp_6, family = binomial,
##     data = data)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.168177   0.014655 -11.476  < 2e-16 ***
## total_imp_1_5  0.032150   0.001464  21.958  < 2e-16 ***
## imp_6          0.014387   0.002929   4.911 9.05e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 34186  on 25300  degrees of freedom
## AIC: 34192
##
## Number of Fisher Scoring iterations: 5
```

Findings:

- Sites 1-5 have a stronger impact on purchases (coefficient: 0.0329031, $p < 2e\text{-}16$).
- Site 6 is effective but less impactful than Sites 1-5 (coefficient: 0.014387, $p < 9.05e\text{-}07$).

Cost & ROI Comparison:

```
total_impressions_1_5 <- sum(data$total_imp_1_5, na.rm = TRUE)
total_impressions_6 <- sum(data$imp_6, na.rm = TRUE)

cat("\nTotal Impressions for Sites 1-5:", total_impressions_1_5, "\n")
```

```
##
## Total Impressions for Sites 1-5: 154144
```

```
cat("Total Impressions for Site 6:", total_impressions_6, "\n")
```

```
## Total Impressions for Site 6: 45127
```

```
cost_per_1000_1_5 <- 25
cost_per_1000_6 <- 20

total_cost_1_5 <- (total_impressions_1_5 / 1000) * cost_per_1000_1_5
total_cost_6 <- (total_impressions_6 / 1000) * cost_per_1000_6

cat("\nTotal Cost for Sites 1-5: $", round(total_cost_1_5, 2), "\n")
```

```
##
## Total Cost for Sites 1-5: $ 3853.6
```

```
cat("Total Cost for Site 6: $", round(total_cost_6, 2), "\n")
```

```
## Total Cost for Site 6: $ 902.54
```

```
total_conversions_1_5 <- sum(data$purchase[data$total_imp_1_5 > 0], na.rm = TRUE)
total_conversions_6 <- sum(data$purchase[data$imp_6 > 0], na.rm = TRUE)

cat("\nTotal Conversions for Sites 1-5:", total_conversions_1_5, "\n")
```

```
##
## Total Conversions for Sites 1-5: 9279
```

```
cat("Total Conversions for Site 6:", total_conversions_6, "\n")
```

```
## Total Conversions for Site 6: 6378
```

```r
cost_per_conversion_1_5 <- ifelse(total_conversions_1_5 > 0, total_cost_1_5 / total_conv
cost_per_conversion_6 <- ifelse(total_conversions_6 > 0, total_cost_6 / total_conversion

cat("\nCost per Conversion for Sites 1-5: $", round(cost_per_conversion_1_5, 2), "\n")
```

```
##
## Cost per Conversion for Sites 1-5: $ 0.42
```

```r
cat("Cost per Conversion for Site 6: $", round(cost_per_conversion_6, 2), "\n")
```

```
## Cost per Conversion for Site 6: $ 0.14
```

```r
revenue_per_conversion <- 1200

total_revenue_1_5 <- total_conversions_1_5 * revenue_per_conversion
total_revenue_6 <- total_conversions_6 * revenue_per_conversion

cat("\nTotal Revenue from Conversions for Sites 1-5: $", total_revenue_1_5, "\n")
```

```
##
## Total Revenue from Conversions for Sites 1-5: $ 11134800
```

```r
cat("Total Revenue from Conversions for Site 6: $", total_revenue_6, "\n")
```

```
## Total Revenue from Conversions for Site 6: $ 7653600
```

```r
roi_1_5 <- ifelse(total_cost_1_5 > 0, (total_revenue_1_5 - total_cost_1_5) / total_cost_
roi_6 <- ifelse(total_cost_6 > 0, (total_revenue_6 - total_cost_6) / total_cost_6, NA)

cat("\nROI for Sites 1-5: ", round(roi_1_5 * 100, 2), "%\n")
```

```
##
## ROI for Sites 1-5:  288845.4 %
```

```r
cat("ROI for Site 6: ", round(roi_6 * 100, 2), "%\n")
```

```
## ROI for Site 6:  847906.7 %
```

Findings:

- Site 6 is 66.67% cheaper(CPC) than Sites 1-5. It also has higher ROI.
- However, Sites 1-5 drive more total conversions and have higher total revenue.
- We would recommend Site 6 as it would give the most returns for the same amount of cost.

Conclusion and Final Recommendations

This analysis provides a **comprehensive evaluation of Star Digital's display advertising strategy**. Our findings suggest that **advertising positively impacts conversions**, but its **effectiveness depends on ad frequency and site selection**. The key takeaways from this study include:

- **Ad exposure significantly increases purchase probability, but there are diminishing returns.**
- **Site selection plays a crucial role**—Sites 1-5 collectively generate **more total conversions**, but **Site 6 is significantly more cost-efficient**.
- **Site 6 offers a lower cost per conversion ($0.14) and the highest ROI (847,906.7%)**, making it the **most efficient option in terms of ad spend.**
- **Future experiments should consider refining targeting strategies** to optimize exposure across sites without excessive overlap.

Strategic Recommendations for Star Digital

Based on these findings, we recommend:

- **Prioritizing Site 6 for advertising** due to **its lower cost per conversion and higher return on investment (ROI).**
- **Shifting budget allocation from Sites 1-5 to Site 6**.
- **Capping impressions per user** to avoid diminishing returns and optimize ad spending.
- **Conducting further experiments** to track **cross-device behavior** and assess if reallocating spend from Sites 1-5 leads to a sustained improvement in overall performance.

By implementing these insights, Star Digital can **maximize its return on advertising spend, reduce acquisition costs, and refine its long-term digital marketing strategy with confidence.**