Code ▾

A/B Testing Analysis of Digital Marketing Campaign

Dang Le (https://www.linkedin.com/in/hdang-le3107/)

May 5, 2025

1. Introduction

This report presents a comprehensive analysis of an A/B testing campaign comparing the effectiveness of advertisements versus public service announcements (PSAs). The analysis examines conversion rates, optimal timing for ad delivery, and factors affecting user engagement.

1.1 Source Information

- **Dataset Title**: Marketing A/B Testing Dataset
- **URL**: Marketing A/B Testing Dataset (https://www.kaggle.com/datasets/faviovaz/marketing-ab-testing/)
- **Who/What is Measured**: Individual users who were exposed either to an advertisement or a public service announcement (PSA) as part of an A/B marketing experiment. The data tracks their engagement and purchasing behavior.

This dataset captures the outcomes of an A/B test run by marketing companies to evaluate whether advertising exposure influences purchasing behavior, and how exposure patterns vary among users.

1.2 Variable Descriptions

Variable Name	Type	What it Measures	Units/Categories
Index	Quantitative	The row index in the dataset	Integer (row number)
User ID	Categorical	A unique identifier assigned to each user	Unique user IDs (alphanumeric)
Test Group	Categorical	Type of content exposed to user	`ad` (advertisement), `psa` (PSA)
Converted	Categorical	Whether the user purchased the product after exposure	`TRUE` (purchased), `FALSE` (not purchased)
Total Ads	Quantitative	Total number of ads seen by each user	Count (number of ads)
Most Ads Day	Categorical	Day of the week user saw the most ads	Days of the week (e.g., Monday, Tuesday)
Most Ads Hour	Quantitative	Hour of the day user saw the most ads	Hour (integer, 24-hour clock)

2. Data Preparation

Hide

```
# Load the marketing_AB dataset
df <- read.csv("marketing_AB.csv")

# Display the first few rows of the dataset
head(df)
```

```
##   X user.id test.group converted total.ads most.ads.day most.ads.hour
## 1 0 1069124         ad     False       130       Monday            20
## 2 1 1119715         ad     False        93      Tuesday            22
## 3 2 1144181         ad     False        21      Tuesday            18
## 4 3 1435133         ad     False       355      Tuesday            10
## 5 4 1015700         ad     False       276       Friday            14
## 6 5 1137664         ad     False       734     Saturday            10
```

2.1 Data Transformation

Hide

```r
# Convert the 'converted' column to logical
df$converted <- as.logical(df$converted)

# Create new most.ads.time variable
df <- df %>%
  mutate(most.ads.time = case_when(
    most.ads.hour >= 0  & most.ads.hour < 6  ~ "Night",
    most.ads.hour >= 6  & most.ads.hour < 12 ~ "Morning",
    most.ads.hour >= 12 & most.ads.hour < 18 ~ "Afternoon",
    most.ads.hour >= 18 & most.ads.hour < 24 ~ "Evening"
  ))

# Convert categorical variables to factors
df$most.ads.day <- as.factor(df$most.ads.day)
df$test.group <- as.factor(df$test.group)
df$most.ads.time <- as.factor(df$most.ads.time)

# Confirm the transformed data
head(df)
```

```
##   X user.id test.group converted total.ads most.ads.day most.ads.hour
## 1 0 1069124         ad     FALSE       130       Monday            20
## 2 1 1119715         ad     FALSE        93      Tuesday            22
## 3 2 1144181         ad     FALSE        21      Tuesday            18
## 4 3 1435133         ad     FALSE       355      Tuesday            10
## 5 4 1015700         ad     FALSE       276       Friday            14
## 6 5 1137664         ad     FALSE       734     Saturday            10
##   most.ads.time
## 1       Evening
## 2       Evening
## 3       Evening
## 4       Morning
## 5     Afternoon
## 6       Morning
```

3. Exploratory Data Analysis

3.1 Data Overview

Hide

```r
# Basic data overview
str(df)
```

```
## 'data.frame':    588101 obs. of  8 variables:
##  $ X            : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ user.id      : int  1069124 1119715 1144181 1435133 1015700 1137664 1116205 14
96843 1448851 1446284 ...
##  $ test.group   : Factor w/ 2 levels "ad","psa": 1 1 1 1 1 1 1 1 1 1 ...
##  $ converted    : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ total.ads    : int  130 93 21 355 276 734 264 17 21 142 ...
##  $ most.ads.day : Factor w/ 7 levels "Friday","Monday",..: 2 6 6 6 1 3 7 4 6 2
...
##  $ most.ads.hour: int  20 22 18 10 14 10 13 18 19 14 ...
##  $ most.ads.time: Factor w/ 4 levels "Afternoon","Evening",..: 2 2 2 3 1 3 1 2 2
1 ...
```

Hide

```r
summary(df)
```

```
##        X             user.id        test.group   converted      
##  Min.   :     0   Min.   : 900000   ad :564577   Mode :logical  
##  1st Qu.:147025   1st Qu.:1143190   psa: 23524   FALSE:573258   
##  Median :294050   Median :1313725                TRUE :14843    
##  Mean   :294050   Mean   :1310692                               
##  3rd Qu.:441075   3rd Qu.:1484088                               
##  Max.   :588100   Max.   :1654483                               
##                                                                 
##    total.ads          most.ads.day   most.ads.hour     most.ads.time   
##  Min.   :   1.00   Friday   :92608   Min.   : 0.00   Afternoon:257839  
##  1st Qu.:   4.00   Monday   :87073   1st Qu.:11.00   Evening  :168172  
##  Median :  13.00   Saturday :81660   Median :14.00   Morning  :142253  
##  Mean   :  24.82   Sunday   :85391   Mean   :14.47   Night    : 19837  
##  3rd Qu.:  27.00   Thursday :82982   3rd Qu.:18.00                     
##  Max.   :2065.00   Tuesday  :77479   Max.   :23.00                     
##                    Wednesday:80908                                     
```

Hide

```r
# Check for missing values
missing_values <- colSums(is.na(df))
print("Missing values per column:")
```

```
## [1] "Missing values per column:"
```

Hide

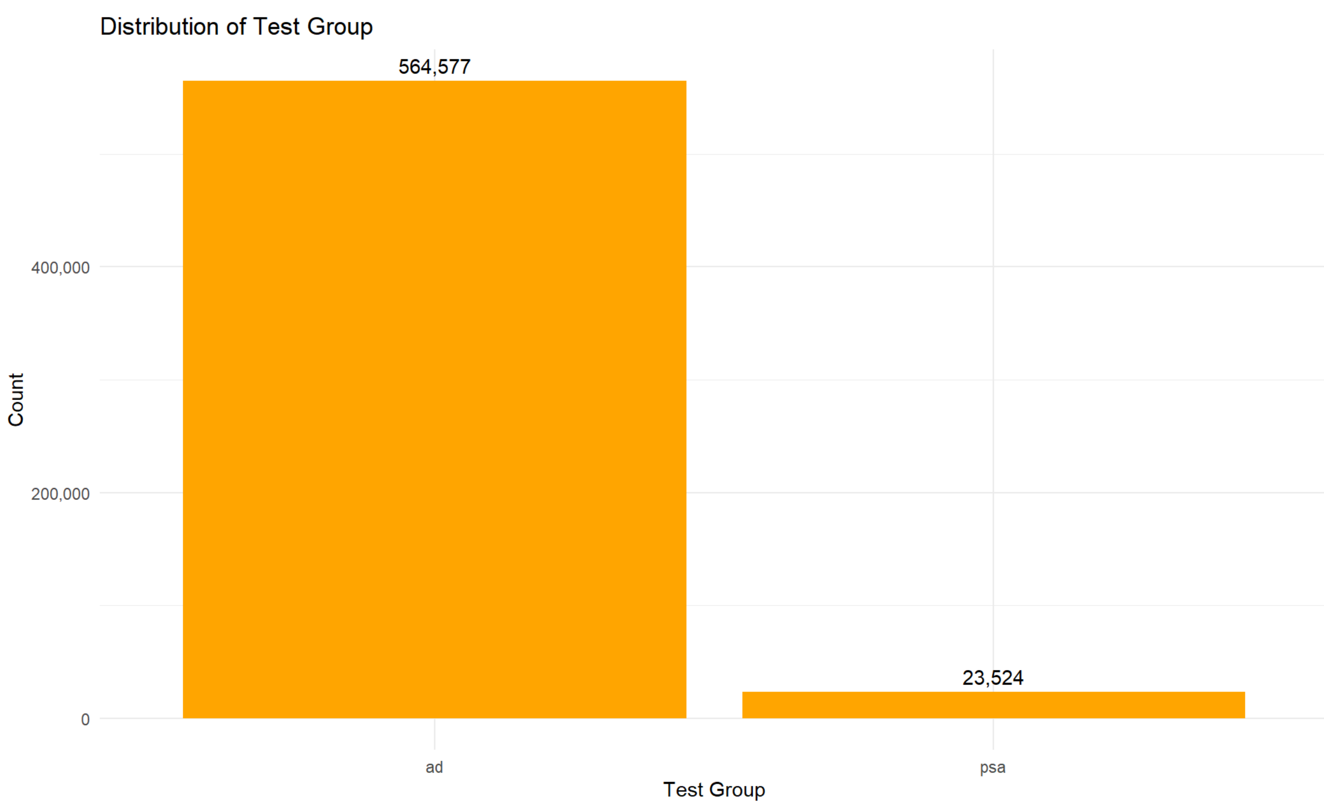
```r
print(missing_values)
```

```
##              X        user.id     test.group      converted      total.ads
##              0              0              0              0              0
##   most.ads.day most.ads.hour most.ads.time
##              0              0              0
```

3.2 Data Distributions

Hide

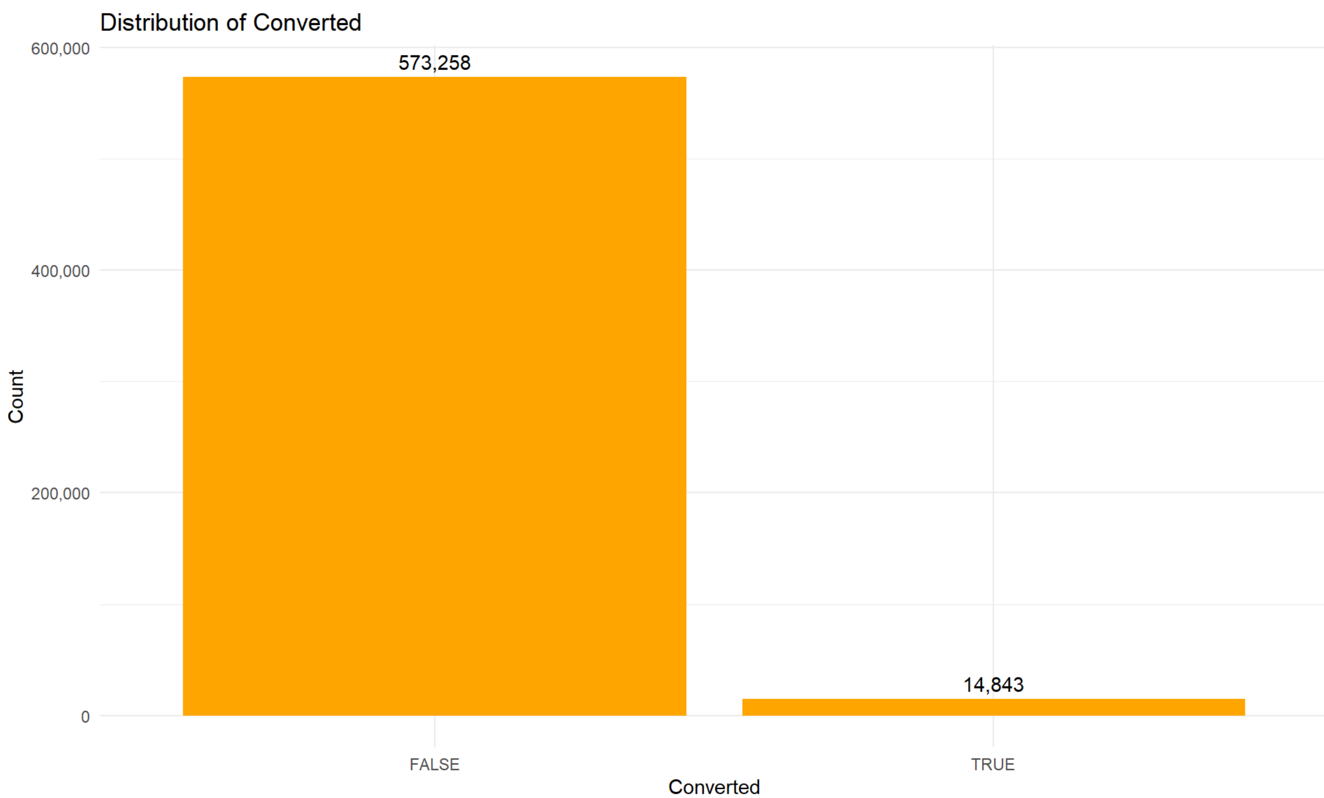
```
# Distribution of test.group
ggplot(df, aes(x = test.group)) +
  geom_bar(fill = "orange") +
  geom_text(stat = "count", aes(label = comma(..count..)), vjust = -0.5) +
  labs(title = "Distribution of Test Group",
       x = "Test Group",
       y = "Count") +
  scale_y_continuous(labels = comma) +
  theme_minimal()
```



Distribution of test groups

Hide

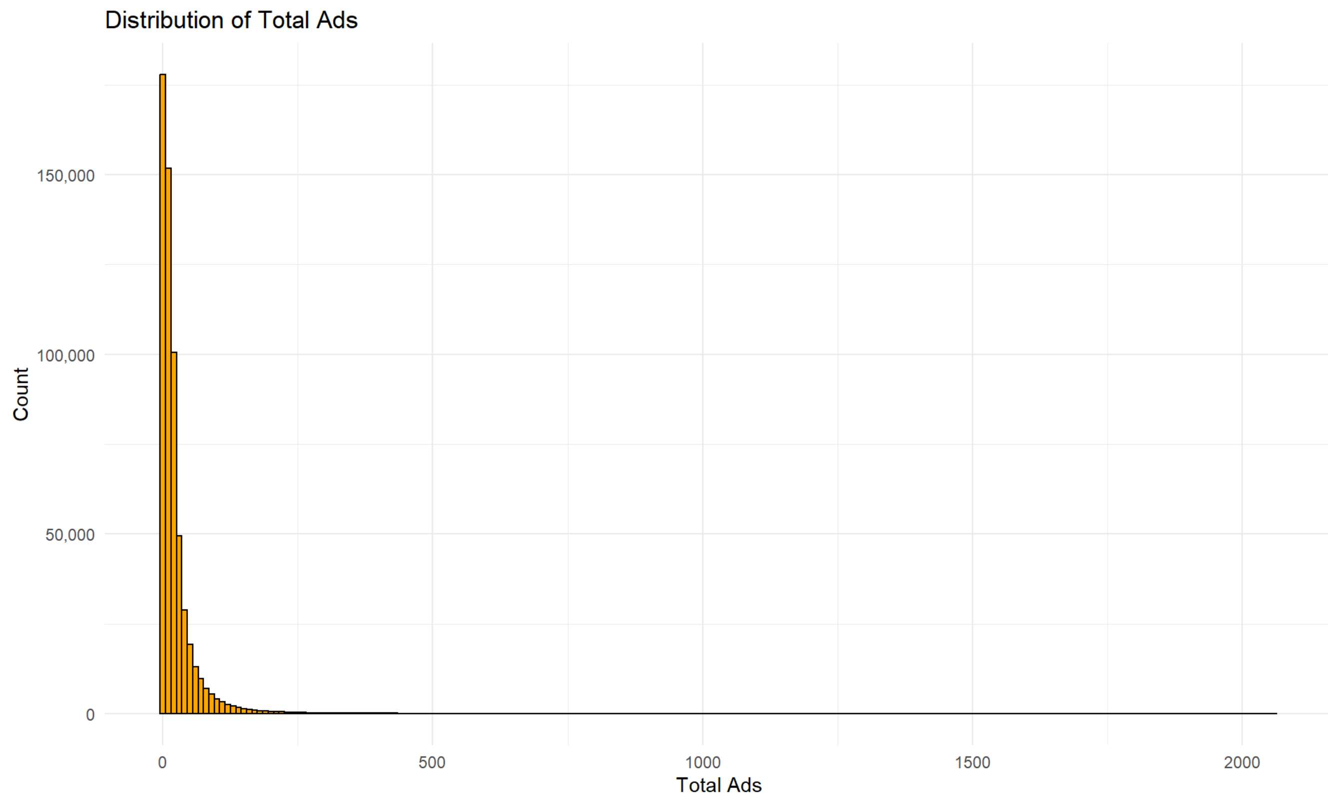
```r
# Distribution of converted
ggplot(df, aes(x = converted)) +
  geom_bar(fill = "orange") +
  geom_text(stat = "count", aes(label = comma(..count..)), vjust = -0.5) +
  labs(title = "Distribution of Converted",
       x = "Converted",
       y = "Count") +
  scale_y_continuous(labels = comma) +
  theme_minimal()
```



Distribution of conversion status

Hide

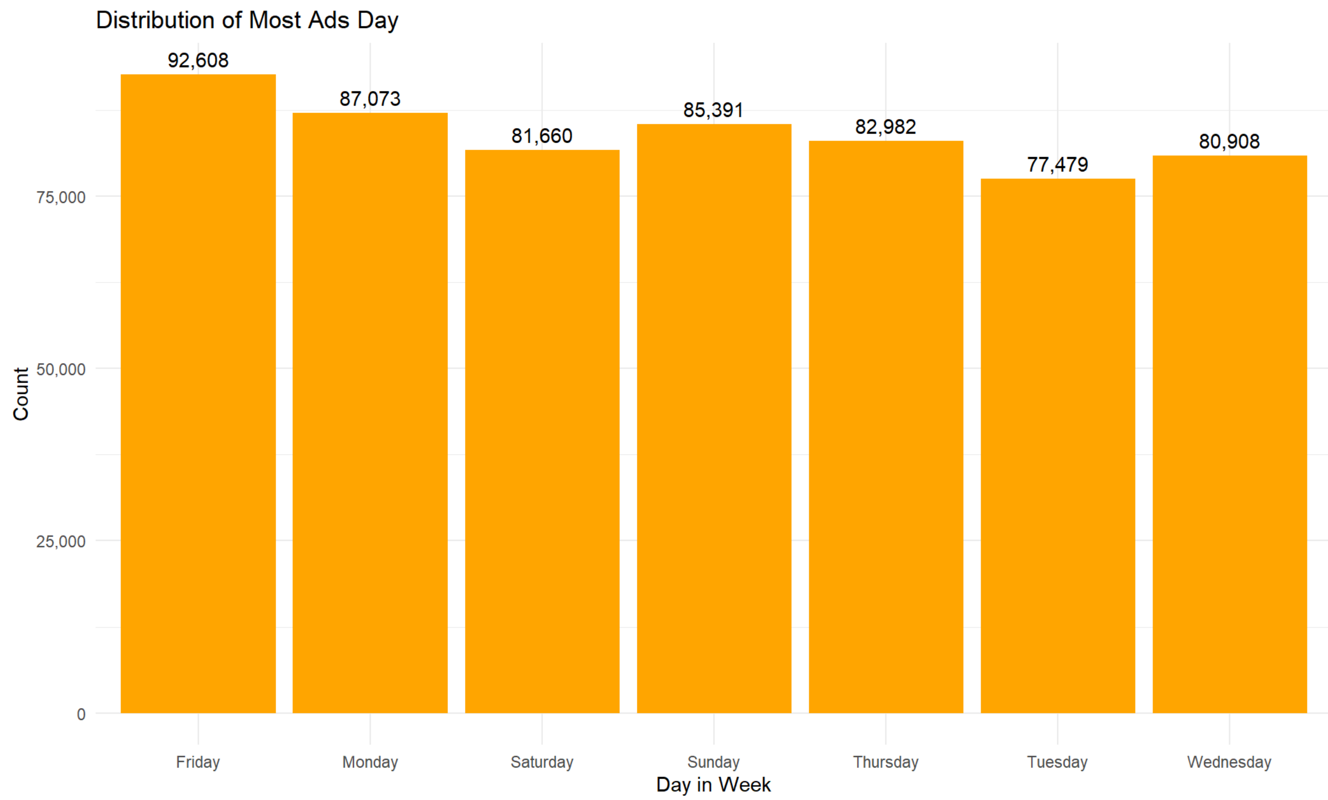
```r
# Distribution of total.ads
ggplot(df, aes(x = total.ads)) +
  geom_histogram(binwidth = 10, fill = "orange", color = "black") +
  labs(title = "Distribution of Total Ads",
       x = "Total Ads",
       y = "Count") +
  scale_y_continuous(labels = comma) +
  theme_minimal()
```



Distribution of total ads

Hide

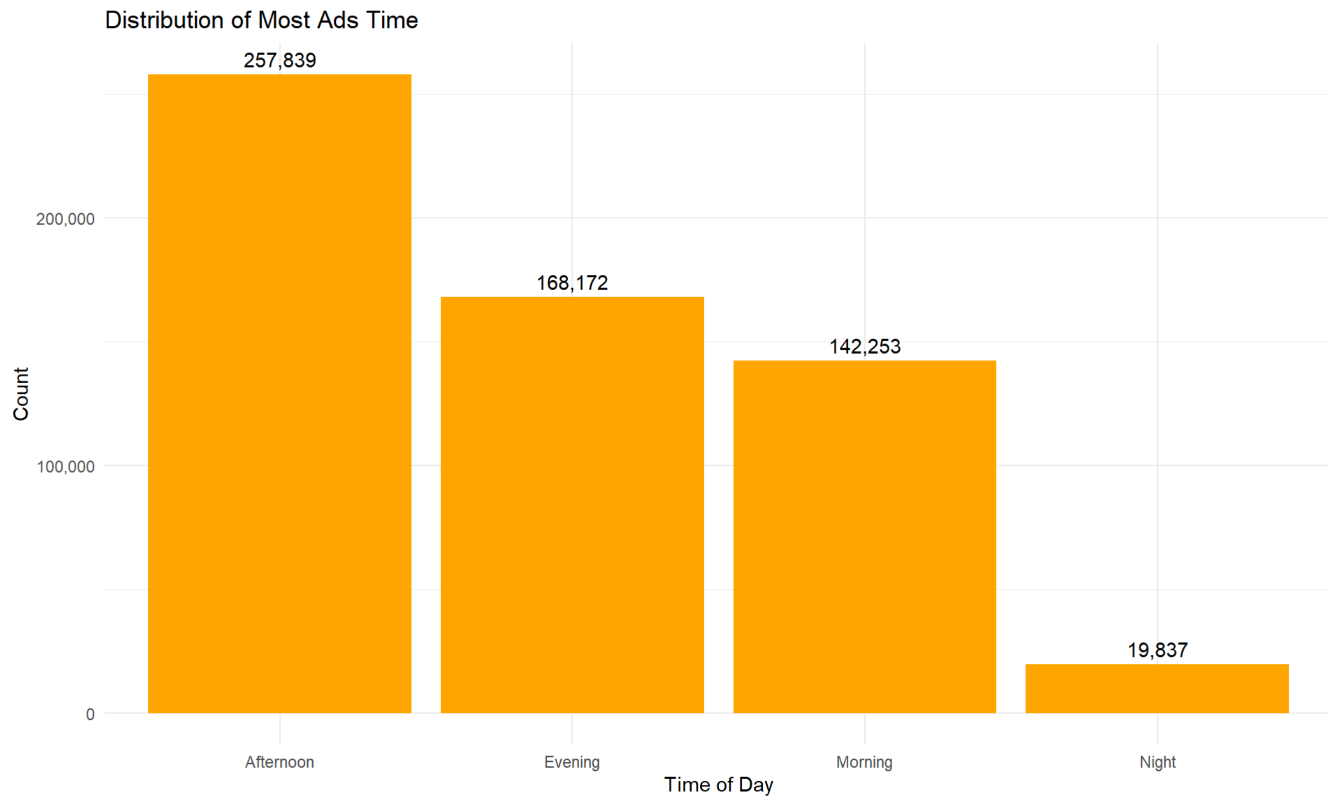
```
# Distribution of most.ads.day
ggplot(df, aes(x = most.ads.day)) +
  geom_bar(fill = "orange") +
  geom_text(stat = "count", aes(label = comma(..count..)), vjust = -0.5) +
  labs(title = "Distribution of Most Ads Day",
       x  = "Day in Week",
       y = "Count") +
  scale_y_continuous(labels = comma) +
  theme_minimal()
```



Distribution of most ads day

Hide

```
# Distribution of most.ads.time
ggplot(df, aes(x = most.ads.time)) +
  geom_bar(fill = "orange") +
  geom_text(stat = "count", aes(label = comma(..count..)), vjust = -0.5) +
  labs(title = "Distribution of Most Ads Time",
       x = "Time of Day",
       y = "Count") +
  scale_y_continuous(labels = comma) +
  theme_minimal()
```

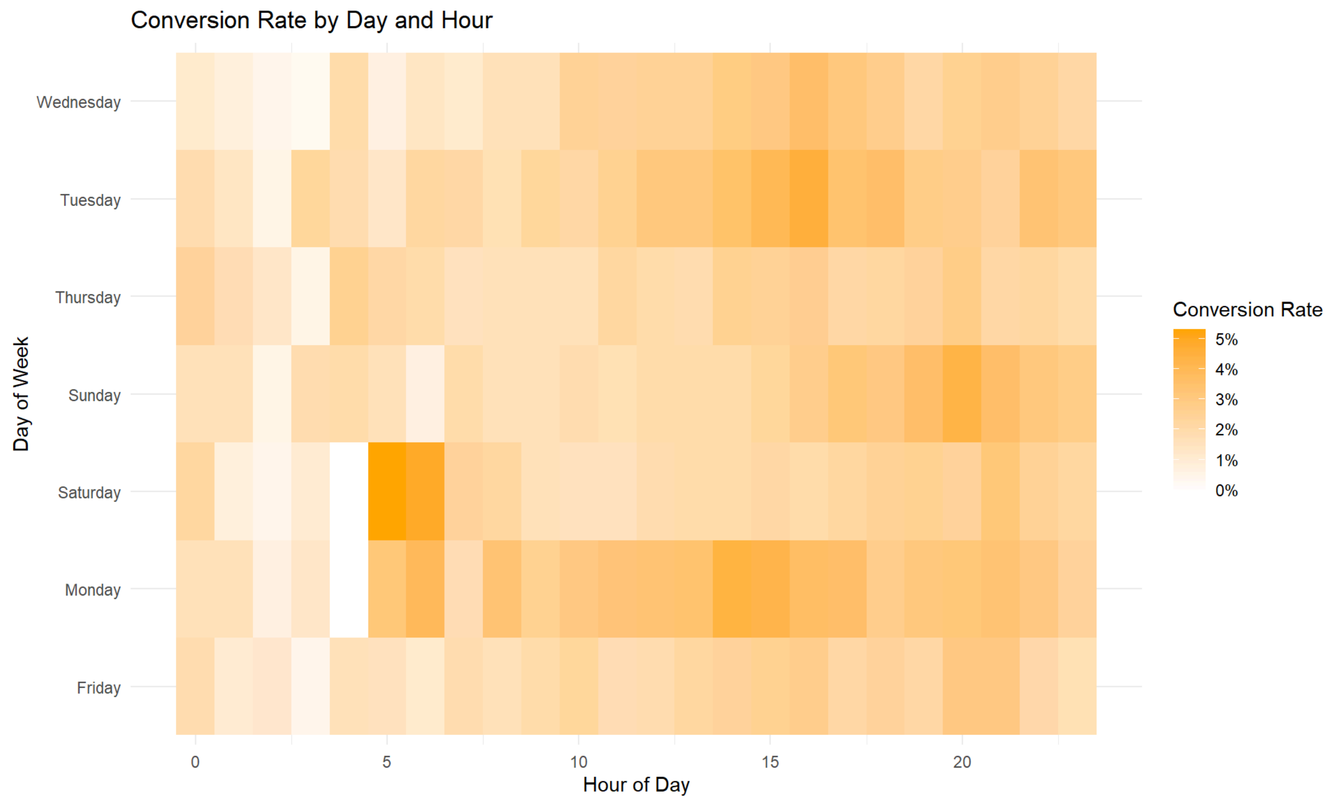


Distribution of most ads time

3.3 Conversion Analysis by Day and Hour

Hide

```
# Heatmap of conversion rate by day and hour
conversion_by_day_hour <- aggregate(
  converted ~ most.ads.day + most.ads.hour,
  data = df,
  FUN = mean
)

# Then plot
ggplot(conversion_by_day_hour, aes(x = most.ads.hour, y = most.ads.day, fill = conve
         rted)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "orange",
                      labels = scales::percent_format()) +
  labs(title = "Conversion Rate by Day and Hour",
       x = "Hour of Day",
       y = "Day of Week",
       fill = "Conversion Rate") +
  theme_minimal()
```



Heatmap of conversion rates by day and hour

4. Power Analysis and Sample Size Verification

Hide

```
# Get current conversion rates from the data
ad_conv_rate <- mean(df$converted[df$test.group == "ad"])
psa_conv_rate <- mean(df$converted[df$test.group == "psa"])
print(paste("Current conversion rates - Ad:", round(ad_conv_rate*100, 2),
            "%, PSA:", round(psa_conv_rate*100, 2), "%"))
```

```
## [1] "Current conversion rates - Ad: 2.55 %, PSA: 1.79 %"
```

Hide

```
# Effect size (difference in proportions)
effect_size <- abs(ad_conv_rate - psa_conv_rate)
print(paste("Observed effect size:", round(effect_size*100, 2), "%"))
```

```
## [1] "Observed effect size: 0.77 %"
```

Hide

```
# Power analysis
power_analysis <- pwr.2p.test(
  h = ES.h(p1 = ad_conv_rate, p2 = psa_conv_rate),
  sig.level = 0.05,
  power = 0.8
)

# Required sample size per group
print("Required sample size per group:")
```

```
## [1] "Required sample size per group:"
```

Hide

```
print(ceiling(power_analysis$n))
```

```
## [1] 5588
```

Hide

```
# Check if our actual sample sizes are sufficient
actual_sizes <- table(df$test.group)
print("Actual sample sizes:")
```

```
## [1] "Actual sample sizes:"
```

Hide

```
print(actual_sizes)
```

```
##
##     ad    psa
## 564577  23524
```

4.1 Power Analysis Summary

Our sample sizes far exceed the minimum requirement, giving us very high statistical power: - Required sample size per group: 5,588 (for reliable detection of the effect) - Actual sample sizes: 564,577 (ad) and 23,524 (PSA)

5. Testing Assumptions for Proportion Tests

Hide

```r
# Calculate number of successes and failures in each group
ad_success <- sum(df$converted[df$test.group == "ad"])
ad_failure <- sum(!df$converted[df$test.group == "ad"])
psa_success <- sum(df$converted[df$test.group == "psa"])
psa_failure <- sum(!df$converted[df$test.group == "psa"])

# Check if conditions are met (np ≥ 10 and n(1-p) ≥ 10 for both groups)
conditions <- data.frame(
  Group = c("ad", "psa"),
  "n×p ≥ 10" = c(ad_success >= 10, psa_success >= 10),
  "n×(1-p) ≥ 10" = c(ad_failure >= 10, psa_failure >= 10)
)

cat("Checking conditions for normal approximation:\n")
```

```
## Checking conditions for normal approximation:
```

Hide

```r
print(conditions)
```

```
##   Group n.p...10 n..1.p....10
## 1    ad     TRUE         TRUE
## 2   psa     TRUE         TRUE
```

6. Hypothesis Testing: Ad vs. PSA

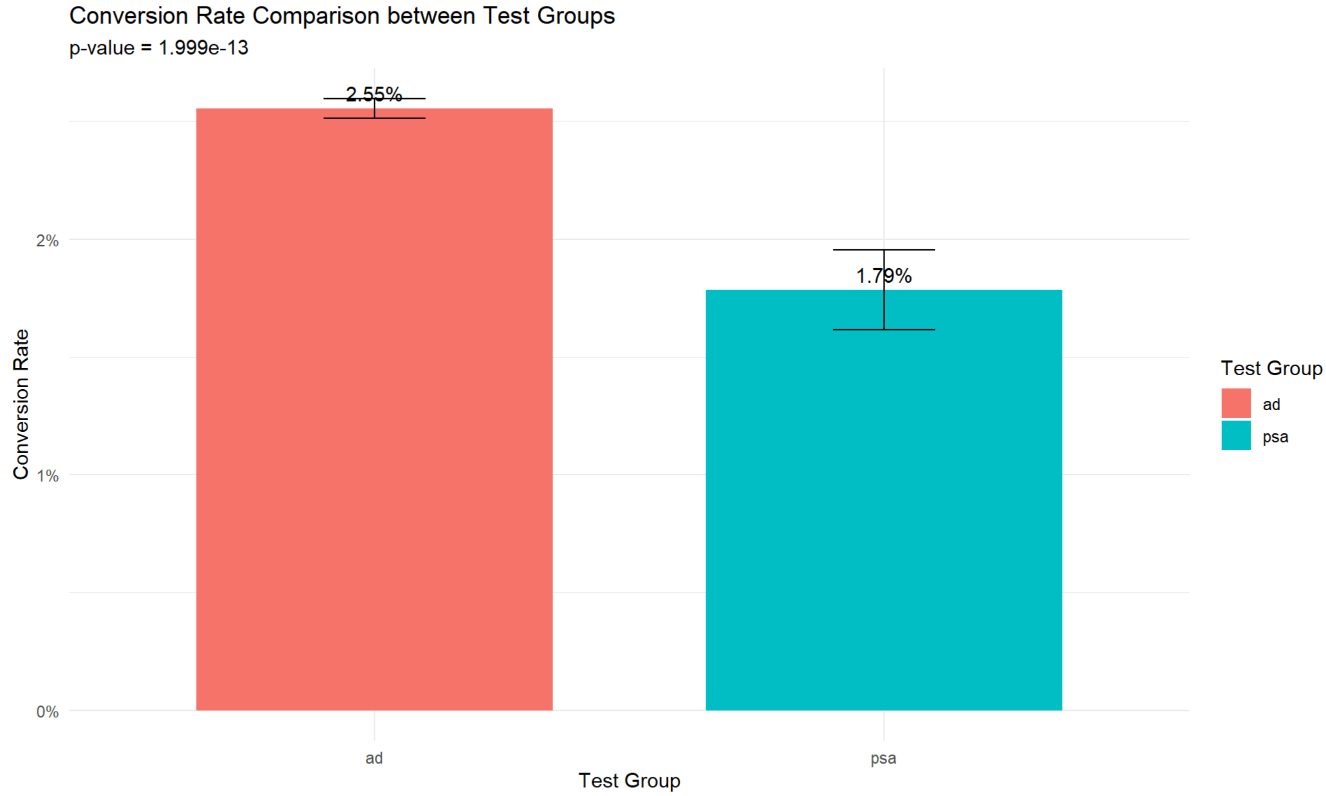
Hide

```r
# Define the groups
ad_group <- df[df$test.group == "ad",]
psa_group <- df[df$test.group == "psa",]

# Perform proportion test
prop_test <- prop.test(
  x = c(sum(ad_group$converted), sum(psa_group$converted)), # Success Convert
  n = c(nrow(ad_group), nrow(psa_group)) # Total Trials
)

prop_test
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(sum(ad_group$converted), sum(psa_group$converted)) out of c(nrow(ad_grou
p), nrow(psa_group))
## X-squared = 54.006, df = 1, p-value = 1.999e-13
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.005928792 0.009456114
## sample estimates:
##     prop 1     prop 2
## 0.02554656 0.01785411
```

Hide

```
# Visualize the results
conversion_data <- data.frame(
  Group = c("ad", "psa"),
  ConversionRate = c(ad_success/(ad_success + ad_failure),
                     psa_success/(psa_success + psa_failure)),
  Count = c(ad_success + ad_failure, psa_success + psa_failure)
)

# Create a bar plot with confidence intervals
ggplot(conversion_data, aes(x = Group, y = ConversionRate, fill = Group)) +
  geom_bar(stat = "identity", position = "dodge", width = 0.7) +
  geom_errorbar(aes(
    ymin = ConversionRate - 1.96 * sqrt(ConversionRate * (1 - ConversionRate) / Coun
         t),
    ymax = ConversionRate + 1.96 * sqrt(ConversionRate * (1 - ConversionRate) / Coun
         t)
  ), width = 0.2) +
  geom_text(aes(label = paste0(round(ConversionRate*100, 2), "%")), vjust = -0.5) +
  labs(
    title = "Conversion Rate Comparison between Test Groups",
    subtitle = paste("p-value =", format.pval(prop_test$p.value, digits = 5)),
    x = "Test Group",
    y = "Conversion Rate",
    fill = "Test Group"
  ) +
  scale_y_continuous(labels = percent_format()) +
  theme_minimal()
```



Comparison of conversion rates between test groups

6.1 Interpretation of Ad vs. PSA Test

Test Information: - **Test Used**: Z-test for equality of proportions (via `prop.test()`)

- **Why This Test**: This test is appropriate when comparing conversion rates (proportions) between two independent groups (ad vs. PSA) with large sample sizes.
- **Null Hypothesis (H_0)**: There is no difference in conversion rates between the ad group and PSA group ($p_1 = p_2$).
- **Alternative Hypothesis (H_1)**: There is a difference in conversion rates between the ad group and PSA group ($p_1 \neq p_2$). Where p_1 and p_2 represent the true conversion rates in the ad and PSA populations, respectively.

Key Findings: - **Conversion Rates**: - Ad group: 2.55% conversion rate - PSA group: 1.79% conversion rate - Observed effect size: 0.77 percentage points difference

- **Statistical Significance**:
  - p-value: 1.999e-13 (extremely small)
  - 95% confidence interval: 0.00593 to 0.00946 (difference in proportions)
  - Since the p-value is much less than 0.05 and the confidence interval doesn't include zero, we reject the null hypothesis and conclude that the difference in conversion rates is highly statistically significant
- **Practical Significance**:
  - The ad group has a 42.5% higher relative conversion rate than the PSA group
  - This represents a substantial improvement in conversion performance

7. Effect of Total Ads on Conversion (Ad Group Only)

Next, we analyze whether the number of total ads affects conversion rates within the ad campaign, without the potential confounding factor of including the PSA group.

Hide

```
# Perform Levene's test for homogeneity of variance
levene_result <- leveneTest(total.ads ~ converted, data = ad_group)
print("Levene's test for equality of variances:")
```

```
## [1] "Levene's test for equality of variances:"
```

Hide

```
print(levene_result)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value    Pr(>F)
## group      1    8739 < 2.2e-16 ***
##       564575
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hide

```
# Based on Levene's test result, perform the appropriate t-test
# Since p-value from Levene's test < 0.05, use var.equal = FALSE
var_equal <- levene_result[1,3] >= 0.05
t_test_result <- t.test(total.ads ~ converted, data = ad_group, var.equal = var_equa
         l)
print("Two-sample t-test results:")
```

```
## [1] "Two-sample t-test results:"
```

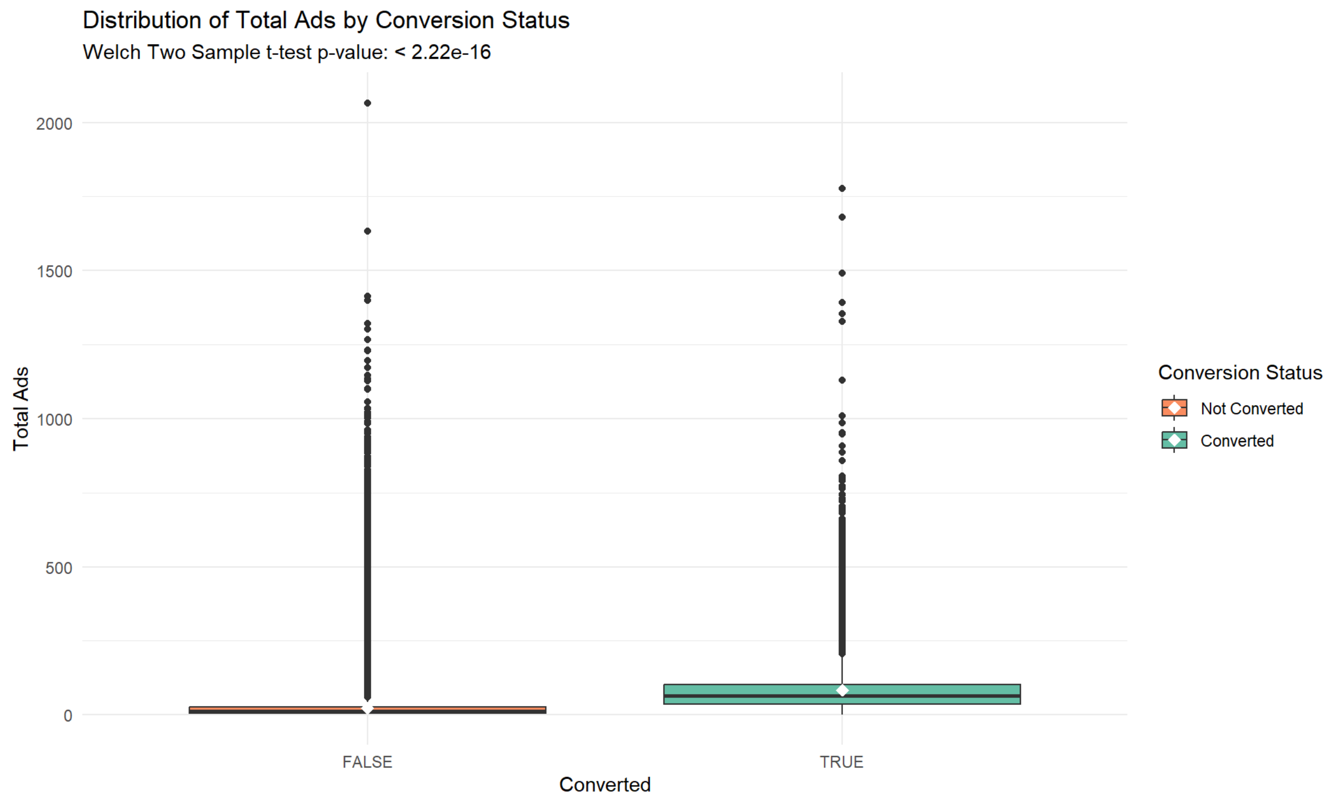
Hide

```
print(t_test_result)
```

```
##
##  Welch Two Sample t-test
##
## data:  total.ads by converted
## t = -82.977, df = 14587, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group FALSE and group TR
UE is not equal to 0
## 95 percent confidence interval:
##  -62.06338 -59.19886
## sample estimates:
## mean in group FALSE  mean in group TRUE
##            23.27445            83.90557
```

Hide

```
# Create a box plot to visualize the differences
ggplot(ad_group, aes(x = converted, y = total.ads, fill = converted)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", shape = 18, size = 3, color = "white") +
  labs(title = "Distribution of Total Ads by Conversion Status",
       subtitle = paste("Welch Two Sample t-test p-value:", format.pval(t_test_resul
         t$p.value, digits = 5)),
       x = "Converted",
       y = "Total Ads",
       fill = "Conversion Status") +
  scale_fill_manual(values = c("FALSE" = "#FC8D62", "TRUE" = "#66C2A5"),
                    labels = c("FALSE" = "Not Converted", "TRUE" = "Converted")) +
  theme_minimal()
```



Distribution of total ads by conversion status

7.1 Interpretation of Total Ads Analysis

Test Information: - **Test Used**: Welch's Two-Sample t-test for difference in means

- **Why This Test**: This test is appropriate for comparing the mean number of ads between converted and non-converted users because: 1. Levene's test showed unequal variances between groups ($p < 0.05$) 2. The sample sizes are large enough to assume approximate normality of sampling distributions 3. The observations are independent

- **Null Hypothesis (H_0)**: There is no difference in the mean number of total ads shown between users who converted and those who did not ($\mu_1 = \mu_2$).

- **Alternative Hypothesis (H_1)**: There is a difference in the mean number of total ads shown between users who converted and those who did not ($\mu_1 \neq \mu_2$). Where μ_1 and μ_2 represent the true mean number of ads shown to converted and non-converted users, respectively.

Statistical Findings: - Test statistic: $t = -82.977$, $df = 14587$ - p-value: $< 2.2e-16$ (extremely small) - 95% confidence interval: -62.06 to -59.2

Key Differences: - Non-converted users were shown an average of 23 ads - Converted users were shown an average of 84 ads - The difference is approximately 61 more ads shown to converted users

Conclusion and Implications: Based on the extremely small p-value, we reject the null hypothesis and conclude that there is a significant difference in the mean number of ads shown to converted versus non-converted users. This strong relationship between ad exposure and conversion suggests increasing ad frequency could potentially improve conversion rates, though correlation doesn't necessarily imply causation.

8. Day of Week Analysis (Ad Group Only)

Here we determine whether there is an association between the day a user sees the most ads and conversion rates.

Hide

```
# Create a contingency table between most.ads.day and converted
day_table <- table(ad_group$converted, ad_group$most.ads.day)
day_conversion_table <- table(ad_group$most.ads.day, ad_group$converted)

# Display the contingency table
print("Contingency table of Most Ads Day vs Conversion (ad group only):")
```

```
## [1] "Contingency table of Most Ads Day vs Conversion (ad group only):"
```

Hide

```
print(day_conversion_table)
```

```
##
##              FALSE  TRUE
##   Friday     86810  1995
##   Monday     80793  2778
##   Saturday   77123  1679
##   Sunday     80305  2027
##   Thursday   77366  1711
##   Tuesday    72302  2270
##   Wednesday 75455  1963
```

Hide

```
# Perform chi-square test for independence
chi_test_day <- chisq.test(day_conversion_table)
print("Chi-square test for independence:")
```

```
## [1] "Chi-square test for independence:"
```

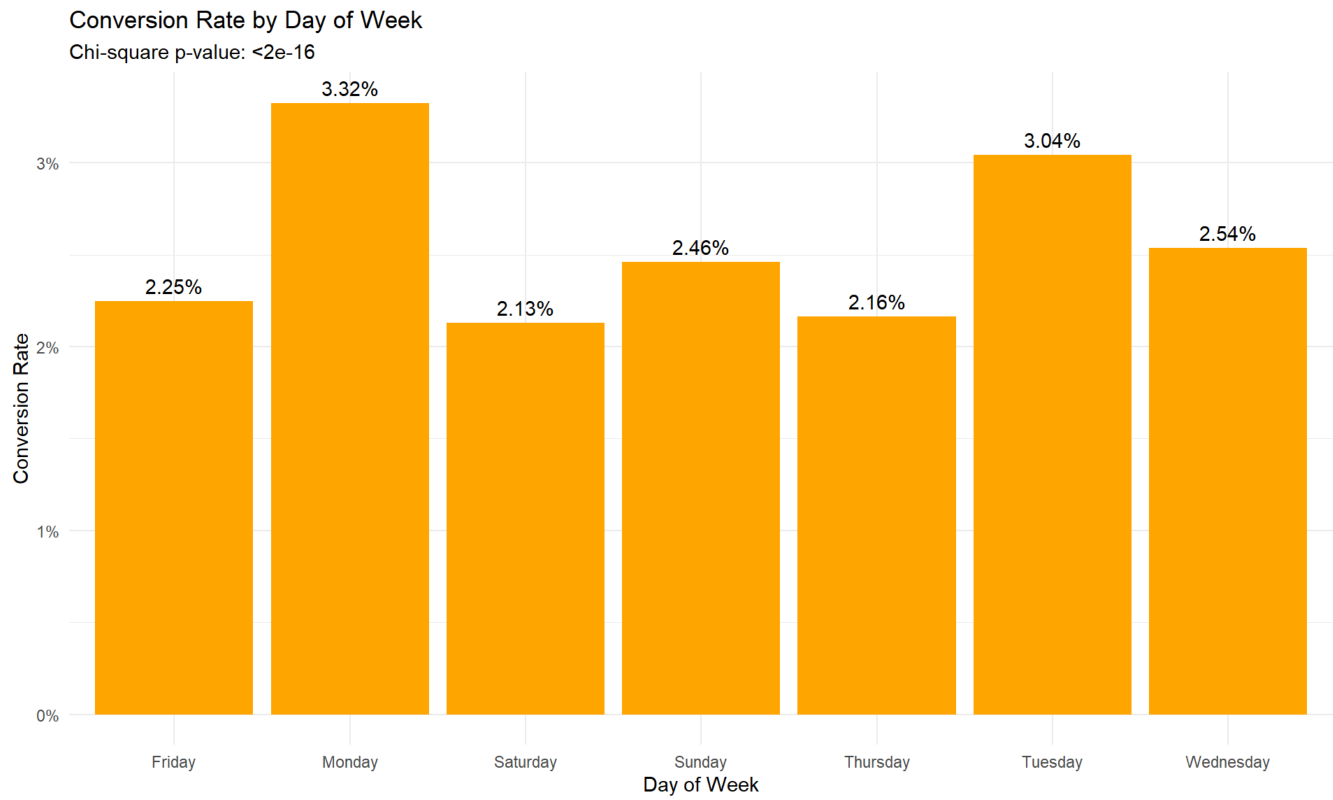
Hide

```
print(chi_test_day)
```

```
##
##  Pearson's Chi-squared test
##
## data:  day_conversion_table
## X-squared = 412.79, df = 6, p-value < 2.2e-16
```

Hide

```
# Calculate conversion rates by day of week
day_conversion_rates <- prop.table(day_table, margin = 2)
conversion_by_day <- data.frame(
  Day = levels(ad_group$most.ads.day),
  ConversionRate = day_conversion_rates["TRUE",]
)

# Create a bar plot showing conversion rates by day of week
ggplot(conversion_by_day, aes(x = Day, y = ConversionRate)) +
  geom_bar(stat = "identity", fill = "orange") +
  geom_text(aes(label = paste0(round(ConversionRate*100, 2), "%")),
            vjust = -0.5) +
  labs(title = "Conversion Rate by Day of Week",
       subtitle = paste("Chi-square p-value:", format.pval(chi_test_day$p.value, dig
         its = 3)),
       x = "Day of Week",
       y = "Conversion Rate") +
  scale_y_continuous(labels = percent_format()) +
  theme_minimal()
```



Conversion rates by day of week

8.1 Interpretation of Day of Week Analysis

Test Information: - **Test Used**: Pearson's Chi-square test for independence

- **Why This Test**: This test is appropriate for examining the relationship between two categorical variables (day of week and conversion status) to determine if conversion rates vary significantly by day.
- **Null Hypothesis (H_0)**: There is no association between day of the week and conversion status. The day of the week when users see most ads is independent of whether they convert (purchase).
- **Alternative Hypothesis (H_1)**: There is an association between day of the week and conversion status. The day of the week when users see most ads is related to their likelihood of converting.

Statistical Findings: - Test statistic: X-squared = 412.79, df = 6 - p-value: < 2.2e-16 (extremely small) - With a p-value far below the conventional significance level of 0.05, we reject the null hypothesis and conclude that there is a significant association between day of week and conversion rates.

Key Insights:

- **Weekday advantage**: Monday (3.32%) and Tuesday (3.04%) show significantly higher conversion rates.
- **Weekend underperformance**: Saturday has the lowest conversion rate (2.13%).
- **Mid-week decline**: Conversion rates generally decline throughout the week.

This analysis provides strong evidence that the day of the week significantly influences conversion rates, with early weekdays performing best.

9. Time of Day Analysis (Ad Group Only)

Here we determine whether there is an association between the time of day a user sees the most ads and conversion rates.

Hide

```
# Create a contingency table between most.ads.time and converted
time_table <- table(ad_group$converted, ad_group$most.ads.time)
time_conversion_table <- table(ad_group$most.ads.time, ad_group$converted)

print("Contingency table of Most Ads Time vs Conversion (ad group only):")
```

```
## [1] "Contingency table of Most Ads Time vs Conversion (ad group only):"
```

Hide

```
print(time_conversion_table)
```

```
##
##               FALSE   TRUE
##   Afternoon 240093   6832
##   Evening   157579   4445
##   Morning   133637   2889
##   Night      18845    257
```

Hide

```
# Perform chi-square test for independence
chi_test_time <- chisq.test(time_conversion_table)
print("Chi-square test for independence:")
```

```
## [1] "Chi-square test for independence:"
```

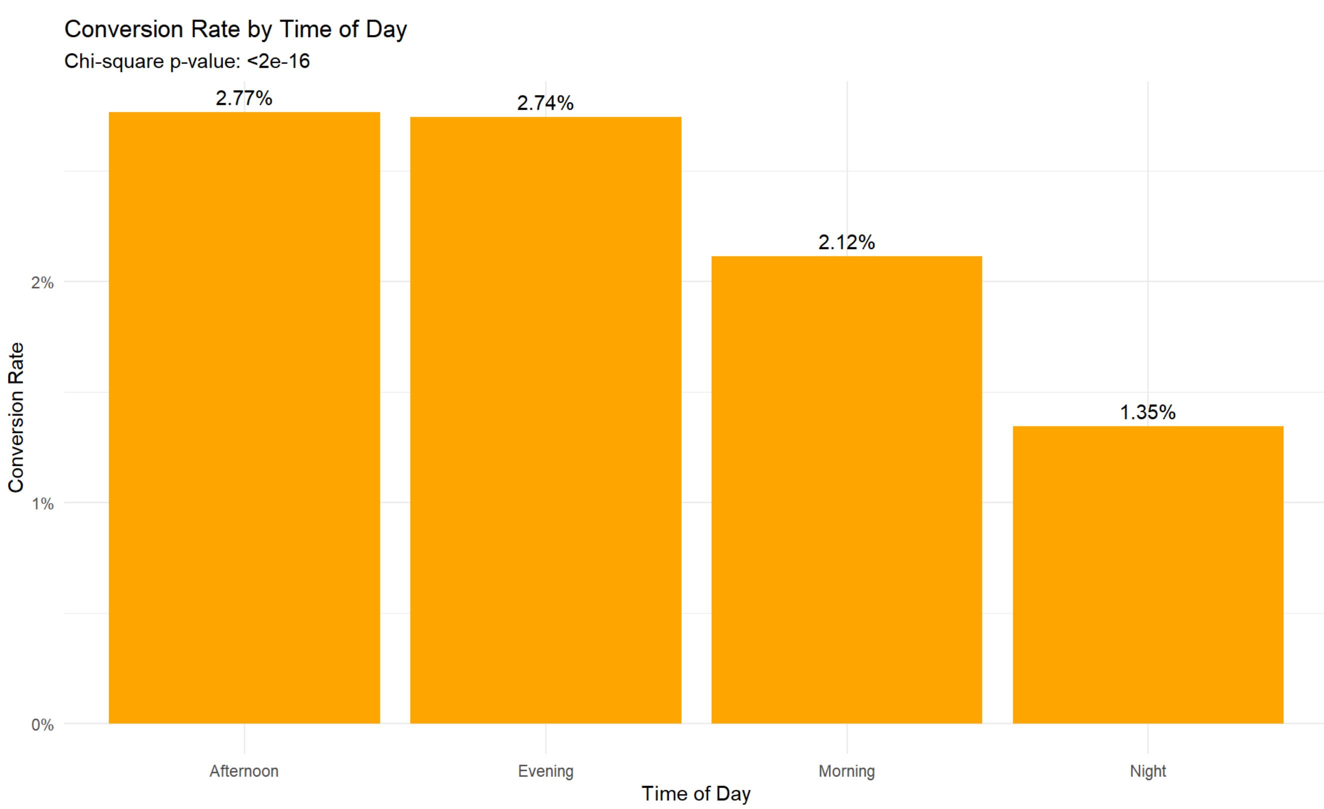
Hide

```
print(chi_test_time)
```

```
##
##  Pearson's Chi-squared test
##
## data:  time_conversion_table
## X-squared = 285.54, df = 3, p-value < 2.2e-16
```

Hide

```r
# Calculate conversion rates by time of day
time_conversion_rates <- prop.table(time_table, margin = 2)
conversion_by_time <- data.frame(
  Time = levels(ad_group$most.ads.time),
  ConversionRate = time_conversion_rates["TRUE",]
)

# Create a bar plot showing conversion rates by time of day
ggplot(conversion_by_time, aes(x = Time, y = ConversionRate)) +
  geom_bar(stat = "identity", fill = "orange") +
  geom_text(aes(label = paste0(round(ConversionRate*100, 2), "%")),
            vjust = -0.5) +
  labs(title = "Conversion Rate by Time of Day",
       subtitle = paste("Chi-square p-value:", format.pval(chi_test_time$p.value, di
gits = 3)),
       x = "Time of Day",
       y = "Conversion Rate") +
  scale_y_continuous(labels = percent_format()) +
  theme_minimal()
```



Conversion rates by time of day

9.1 Interpretation of Time of Day Analysis

Test Information: - Test Used: Pearson’s Chi-square test for independence

- **Why This Test**: This test is appropriate for examining the relationship between two categorical variables (time of day and conversion status) to determine if conversion rates vary significantly by time period.
- **Null Hypothesis (H_0)**: There is no association between time of day and conversion status. The time of day when users see most ads is independent of whether they convert (purchase).
- **Alternative Hypothesis (H_1)**: There is an association between time of day and conversion status. The time of day when users see most ads is related to their likelihood of converting.

Statistical Findings: - Test statistic: X-squared = 285.54, df = 3 - p-value: < 2.2e-16 (extremely small) - With a p-value far below the conventional significance level of 0.05, we reject the null hypothesis and conclude that there is a significant association between time of day and conversion rates.

Key Insights:

- **Afternoon and Evening advantage**: Both afternoon (2.77%) and evening (2.74%) hours show significantly higher conversion rates.
- **Morning underperformance**: Morning has a lower conversion rate (2.12%).
- **Night significant drop**: Night has by far the lowest conversion rate (1.35%).

This analysis provides strong evidence that the time of day significantly influences conversion rates, with afternoon and evening hours performing best.

10. Conclusion and Recommendations

Our comprehensive A/B testing analysis has revealed several important insights that can guide future marketing strategies:

1. **Ad Effectiveness**: The ad campaign significantly outperforms PSAs with a 42.5% higher relative conversion rate (2.55% vs 1.79%).
2. **Ad Frequency**: Users who convert are exposed to substantially more ads (84 ads vs 23 ads). While correlation doesn't prove causation, increasing ad frequency may improve conversion rates.
3. **Optimal Timing - Day of Week**: Monday (3.32%) and Tuesday (3.04%) show the highest conversion rates, while Saturday (2.13%) shows the lowest. Consider reallocating advertising budget to prioritize early weekdays.
4. **Optimal Timing - Time of Day**: Afternoon (2.77%) and evening (2.74%) hours yield significantly higher conversion rates than morning (2.12%) or night (1.35%) hours. Reduce spending during night hours and reallocate to afternoon and evening.
5. **Sample Size Adequacy**: With over 560,000 observations, our sample size far exceeds the minimum required (5,588 per group), giving our findings strong statistical power.

Final Recommendations:

1. Continue using the ad campaign instead of PSAs

2. Increase ad frequency for users showing engagement signals
3. Optimize ad scheduling to prioritize:
   - Days: Monday and Tuesday
   - Times: Afternoon and evening hours
4. Consider conducting follow-up tests to determine the optimal ad frequency threshold that maximizes conversion without causing ad fatigue
5. Develop specialized content for weekend days to address the lower conversion rates

Implementing these recommendations should result in improved conversion rates and more efficient allocation of marketing resources.