

A/B Testing

2024-09-02

Introduction

A/B testing is a method used to compare and evaluate different marketing strategies, designs, or themes. The primary purpose is to make informed decisions based on data to enhance user experience and improve performance metrics. In this project, the aim is to test two themes (Light and Dark) to determine which one looks better on the website.

Dataset

```
# Load necessary libraries
```

```
library(ggplot2)
```

```
library(ggpubr)
```

```
library(dplyr)
```

```
##
```

```
## Attachement du package : 'dplyr'
```

```
## Les objets suivants sont masqués depuis 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## Les objets suivants sont masqués depuis 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
# Importing the dataset
```

```
Data <- read.csv("C:/Users/Mus/Downloads/website_ab_test.csv")
```

```
head(Data)
```

```
##           Theme Click.Through.Rate Conversion.Rate Bounce.Rate Scroll_Depth
Age
```

```
## 1 Light Theme      0.05491998      0.28236683    0.4050849      72.48946
```

```
25
```

```
## 2 Light Theme      0.11393185      0.03297321    0.7327588      61.85857
```

```
19
```

```
## 3 Dark Theme       0.32335153      0.17876327    0.2965427      45.73738
```

```
47
```

```
## 4 Light Theme      0.48583572      0.32522487    0.2450015      76.30530
```

```
58
```

```
## 5 Light Theme      0.03478347      0.19676577    0.7651005      48.92741
```

```
25
```

```
## 6 Light Theme      0.17341949      0.26698690    0.7868249      66.12030
```

```
65
```

```
##      Location Session_Duration Purchases Added_to_Cart
```

```
## 1    Chennai          1535          No              Yes
```

```
## 2      Pune      303      No      Yes
## 3    Chennai      563      Yes     Yes
## 4      Pune      385      Yes     No
## 5 New Delhi     1437      No      No
## 6    Chennai     1244      Yes     No
```

Counting missing values

```
nbr_missing_values <- colSums(is.na(Data))
print(nbr_missing_values)
```

```
##           Theme Click.Through.Rate Conversion.Rate
Bounce.Rate
##           0           0           0
0
##      Scroll_Depth           Age           Location
Session_Duration
##           0           0           0
0
##      Purchases      Added_to_Cart
##           0           0
```

The dataset has no missing values.

Descriptive statistics

```
summary(Data)
```

```
##      Theme      Click.Through.Rate Conversion.Rate      Bounce.Rate
## Length:1000    Min.   :0.01077    Min.   :0.01088    Min.   :0.2007
## Class :character 1st Qu.:0.14079    1st Qu.:0.13156    1st Qu.:0.3536
## Mode  :character Median :0.25372    Median :0.25282    Median :0.5140
##              Mean  :0.25605    Mean  :0.25331    Mean  :0.5058
##              3rd Qu.:0.37067    3rd Qu.:0.37304    3rd Qu.:0.6486
##              Max.   :0.49999    Max.   :0.49892    Max.   :0.7997
##      Scroll_Depth      Age      Location      Session_Duration
## Min.   :20.01    Min.   :18.00    Length:1000    Min.   : 38.0
## 1st Qu.:35.66    1st Qu.:29.00    Class :character 1st Qu.: 466.5
## Median :51.13    Median :42.00    Mode  :character Median : 931.0
## Mean   :50.32    Mean   :41.53              Mean  : 925.0
## 3rd Qu.:64.67    3rd Qu.:54.00              3rd Qu.:1375.2
## Max.   :80.00    Max.   :65.00              Max.   :1797.0
##      Purchases      Added_to_Cart
## Length:1000      Length:1000
## Class :character Class :character
## Mode  :character Mode  :character
##
##
##
```

Visualizations

Create a scatter plot for Click Through Rate (CTR) vs Conversion Rate

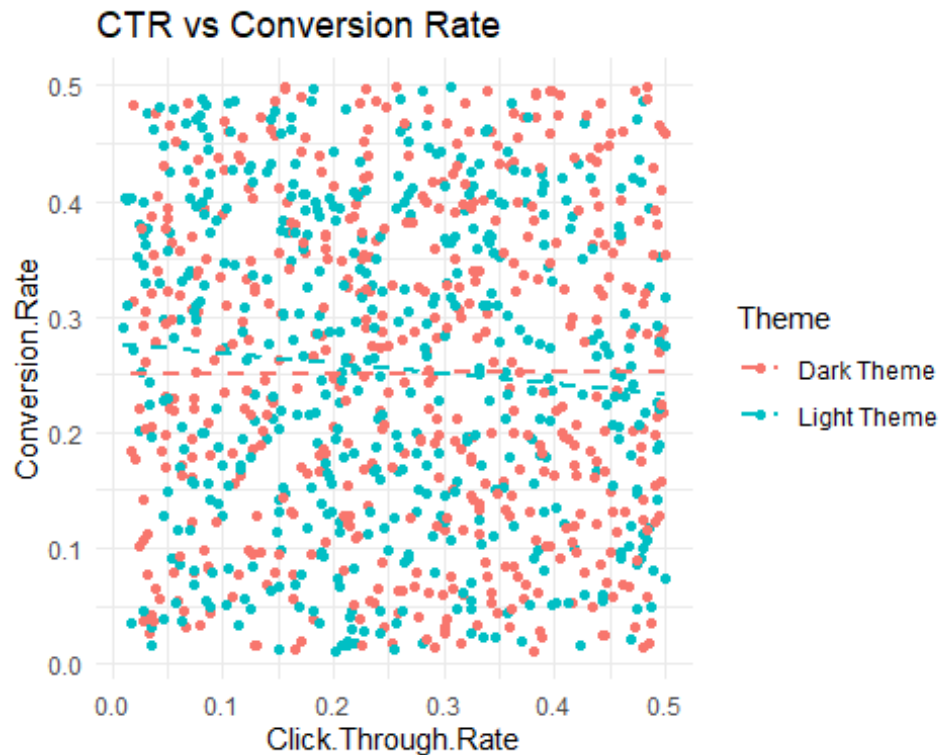
```
ggplot(Data, aes(x = Click.Through.Rate, y = Conversion.Rate, color = Theme))
```

```

+
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, aes(group = Theme), linetype =
"dashed") +
  labs(title = "CTR vs Conversion Rate") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

```



The scatter plot indicates a consistent relationship between the **Click-Through Rate** and the **Conversion Rate**. This suggests that as the number of users clicking on links or buttons increases, a consistent proportion of them proceed to sign up each day.

```

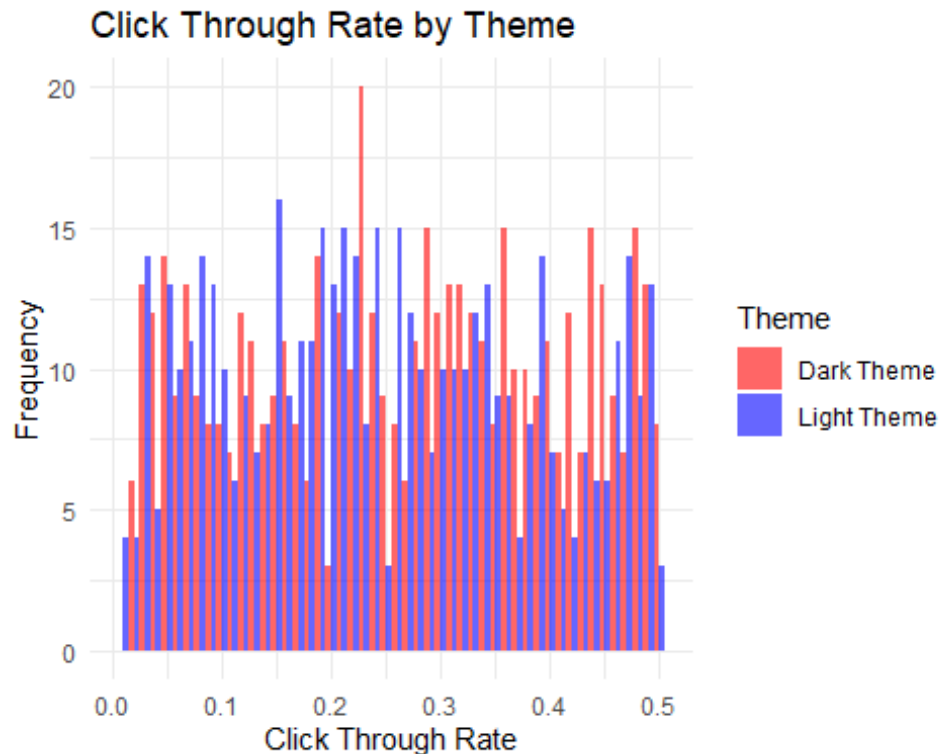
# Extract data for each theme
light_theme_data <- Data %>% filter(Theme == 'Light Theme')
dark_theme_data <- Data %>% filter(Theme == 'Dark Theme')

# Combine the data for the two themes into a single dataframe
combined_data <- bind_rows(
  light_theme_data %>% mutate(Theme = 'Light Theme'),
  dark_theme_data %>% mutate(Theme = 'Dark Theme')
)

# Create a grouped bar chart for Click Through Rate by theme
ggplot(combined_data, aes(x = Click.Through.Rate, fill = Theme)) +
  geom_histogram(position = "dodge", binwidth = 0.01, alpha = 0.6) +
  labs(title = 'Click Through Rate by Theme', x = 'Click Through Rate', y =

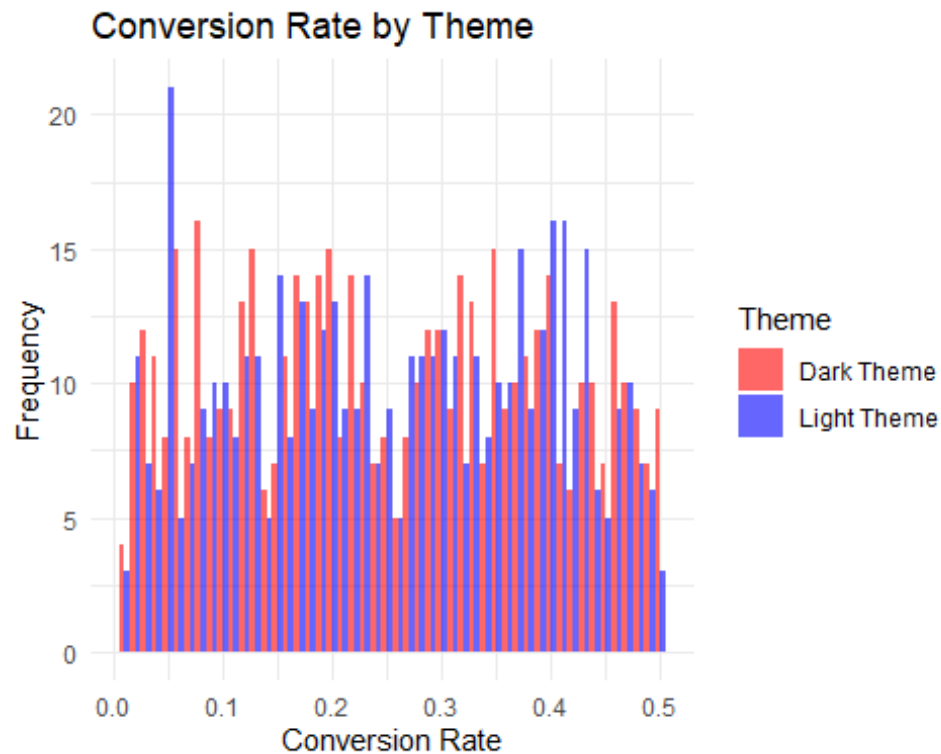
```

```
'Frequency') +
  scale_fill_manual(values = c('Light Theme' = 'blue', 'Dark Theme' = 'red'))
+
  theme_minimal()
```



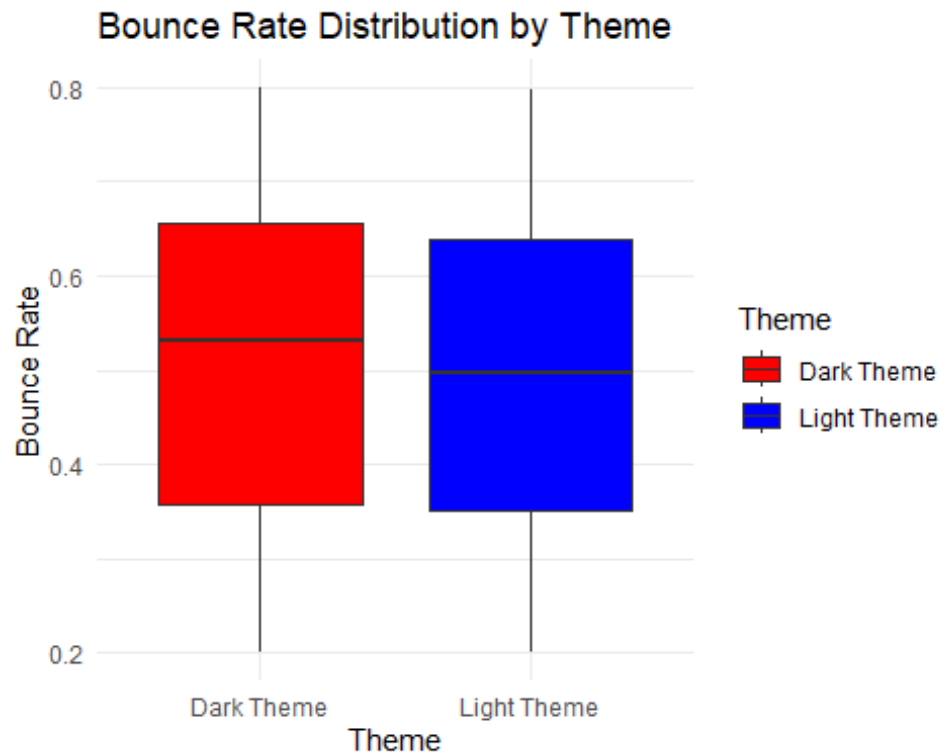
The histogram shows that both themes present the same performance, with the dark theme having a slightly higher click-through rate than the light theme.

```
# Create a grouped bar chart for Conversion Rate by theme
ggplot(combined_data, aes(x = Conversion.Rate, fill = Theme)) +
  geom_histogram(position = "dodge", binwidth = 0.01, alpha = 0.6) +
  labs(title = 'Conversion Rate by Theme', x = 'Conversion Rate', y =
'Frequency') +
  scale_fill_manual(values = c('Light Theme' = 'blue', 'Dark Theme' = 'red'))
+
  theme_minimal()
```



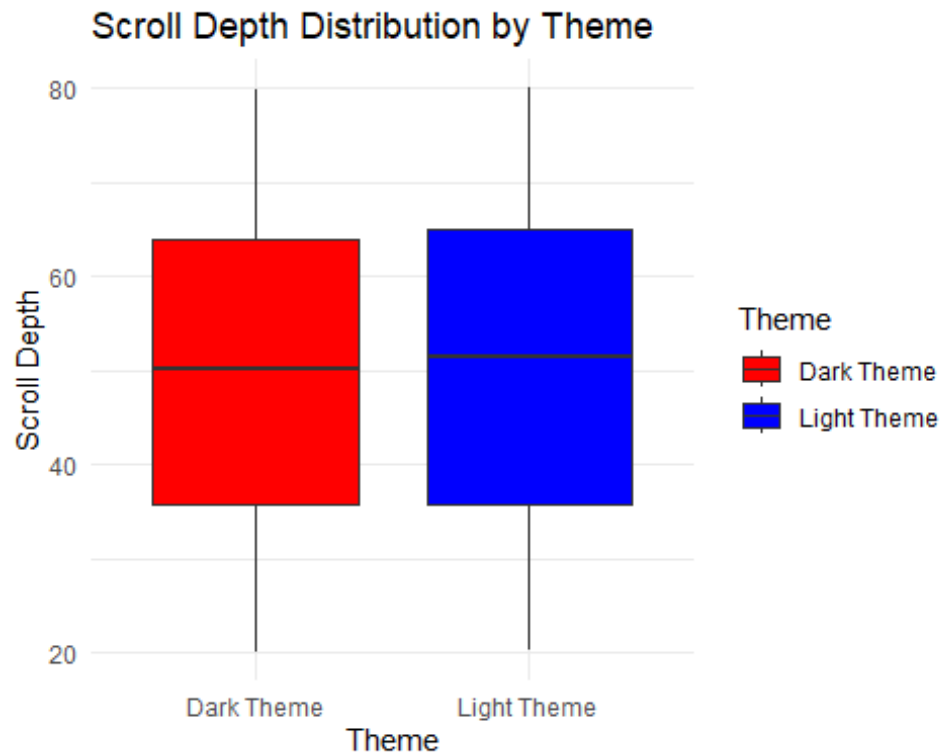
We can see in the histogram that there is minimal difference, with the dark theme having a somewhat better conversion rate compared to the light theme.

```
# Boxplot for Bounce Rate by Theme
ggplot(Data, aes(x = Theme, y = Bounce.Rate, fill = Theme)) +
  geom_boxplot() +
  labs(title = 'Bounce Rate Distribution by Theme', x = 'Theme', y = 'Bounce
Rate') +
  scale_fill_manual(values = c('Light Theme' = 'blue', 'Dark Theme' = 'red'))
+
  theme_minimal()
```



The graph shows that the bounce rates for both themes are nearly identical. The light theme has a marginally lower bounce rate, indicating slightly better performance.

```
# Boxplot for Scroll Depth by Theme
ggplot(Data, aes(x = Theme, y = Scroll_Depth, fill = Theme)) +
  geom_boxplot() +
  labs(title = 'Scroll Depth Distribution by Theme', x = 'Theme', y = 'Scroll
Depth') +
  scale_fill_manual(values = c('Light Theme' = 'blue', 'Dark Theme' = 'red'))
+
  theme_minimal()
```



Concerning the scroll depth, there isn't much difference; but, the light theme performs slightly better.

A/B testing for Purchases

Calculate conversion counts and sample sizes

```
light_theme_conversions <- sum(light_theme_data$Purchases == 'Yes')
light_theme_total <- nrow(light_theme_data)
```

```
dark_theme_conversions <- sum(dark_theme_data$Purchases == 'Yes')
dark_theme_total <- nrow(dark_theme_data)
```

Calculate conversion rates

```
light_theme_conversion_rate <- light_theme_conversions / light_theme_total
dark_theme_conversion_rate <- dark_theme_conversions / dark_theme_total
```

Perform two-sample proportion test

Proportions test for two independent samples

```
test_result <- prop.test(x = c(light_theme_conversions,
                                dark_theme_conversions),
                          n = c(light_theme_total, dark_theme_total))
```

Print conversion rates

```
cat("Light Theme Conversion Rate:", light_theme_conversion_rate, "\n")
```

```
## Light Theme Conversion Rate: 0.5308642
```

```

cat("Dark Theme Conversion Rate:", dark_theme_conversion_rate, "\n")
## Dark Theme Conversion Rate: 0.5038911

# Print test results
cat("A/B Testing - Chi-squared Test Statistic:", test_result$statistic, "\n")
## A/B Testing - Chi-squared Test Statistic: 0.6238119

cat("p-value:", test_result$p.value, "\n")
## p-value: 0.4296343

```

To compare the conversion rates for purchases between the two themes, we conducted an A/B test to see if there is a statistically significant difference.

The positive z-statistic value of 0.62 indicates that the conversion rate of the Light Theme is slightly higher than the conversion rate of the Dark Theme.

The p-value is approximately 0.43, which is greater than 5%. This means we cannot reject the null hypothesis that there is no difference in conversion rates between the two themes. In other words, we cannot conclude that one theme performs better in terms of purchases.

A/B testing for Session Duration

```

# Extract session duration for each theme
light_theme_session_duration <- light_theme_data$Session_Duration
dark_theme_session_duration <- dark_theme_data$Session_Duration

# Calculate the average session duration for both themes
light_theme_avg_duration <- mean(light_theme_session_duration, na.rm = TRUE)
dark_theme_avg_duration <- mean(dark_theme_session_duration, na.rm = TRUE)

# Perform two-sample t-test for session duration
t_test_result <- t.test(light_theme_session_duration,
dark_theme_session_duration)

# Print the average session duration for both themes
cat("Light Theme Average Session Duration:", light_theme_avg_duration, "\n")
## Light Theme Average Session Duration: 930.8333

cat("Dark Theme Average Session Duration:", dark_theme_avg_duration, "\n")
## Dark Theme Average Session Duration: 919.4825

# Print the t-test results
cat("A/B Testing for Session Duration - t-statistic:",
t_test_result$statistic, "\n")
## A/B Testing for Session Duration - t-statistic: 0.3529118

```



```
cat("p-value:", t_test_result$p.value, "\n")  
## p-value: 0.7242293
```

To evaluate the session duration between the two themes, we implemented an A/B test to determine if there is a statistically significant difference.

The positive z-statistic value of 0.35 shows that the average session duration for the Light Theme is somewhat higher than that of the Dark Theme.

The p-value is approximately 0.72, which is greater than 5%. This means we cannot reject the null hypothesis that there is no difference average session duration between the two themes. In simpler terms, the results indicate that the average session duration for both themes is similar.