Portfolio Task 2 - Advertising Experiment Analysis

Narges

2025-02-19

Load and Clean Data

```
# Load dataset
df <- read.csv("/Users/Narges/Downloads/GRA 4158/CaseData2025.csv", sep=";")

# Rename columns
colnames(df) <- c("test", "purchase", "impressions")

# Convert columns to appropriate types
df <- df %>% mutate(
  test = as.factor(test),  # 0 = Control, 1 = Treatment
  purchase = as.integer(purchase),
  impressions = as.integer(impressions)
)

# Display first few rows
head(df)
```

```
##   test purchase impressions
## 1    0        0           1
## 2    1        0           1
## 3    1        0           1
## 4    1        0           2
## 5    1        0           6
## 6    1        0           2
```

Descriptive Statistics

```
# Calculate conversion rates
conversion_rates <- df %>%
  group_by(test) %>%
  summarise(conversion_rate = mean(purchase))

# Baseline conversion rate
baseline_conversion <- conversion_rates$conversion_rate[conversion_rates$test == 0]

# Incremental lift
incremental_lift <- conversion_rates$conversion_rate[conversion_rates$test == 1] - baseline_conversion
```

```r
# Print results
conversion_rates
```

```
## # A tibble: 2 x 2
##   test  conversion_rate
##   <fct>           <dbl>
## 1 0              0.0254
## 2 1              0.0310
```

```r
incremental_lift
```

```
## [1] 0.005586106
```

Statistical Significance Test (Proportion Test)

```r
# Perform a proportion test to check significance of conversion rate difference
prop_test <- prop.test(x = c(sum(df$purchase[df$test == 0]), sum(df$purchase[df$test == 1])),
                       n = c(sum(df$test == 0), sum(df$test == 1)))

prop_test
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(sum(df$purchase[df$test == 0]), sum(df$purchase[df$test == 1])) out of c(sum(df$test == 0),
## X-squared = 3.2105, df = 1, p-value = 0.07317
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.0113443920  0.0001721806
## sample estimates:
##     prop 1     prop 2
## 0.02541296 0.03099907
```

Profitability Analysis

```r
# Given values
revenue_per_purchase <- 300  # NOK per purchase
ad_cost_per_1000_impressions <- 100  # NOK per 1000 impressions

# Total purchases & revenue
total_purchases <- sum(df$purchase)
total_revenue <- total_purchases * revenue_per_purchase

# Total advertising costs
total_impressions <- sum(df %>% filter(test == 1) %>% pull(impressions))
total_ad_cost <- (total_impressions / 1000) * ad_cost_per_1000_impressions
```

```r
# Net profit
net_profit <- total_revenue - total_ad_cost

# Print results
total_revenue
```

```
## [1] 179400
```

```r
total_ad_cost
```

```
## [1] 4564.9
```

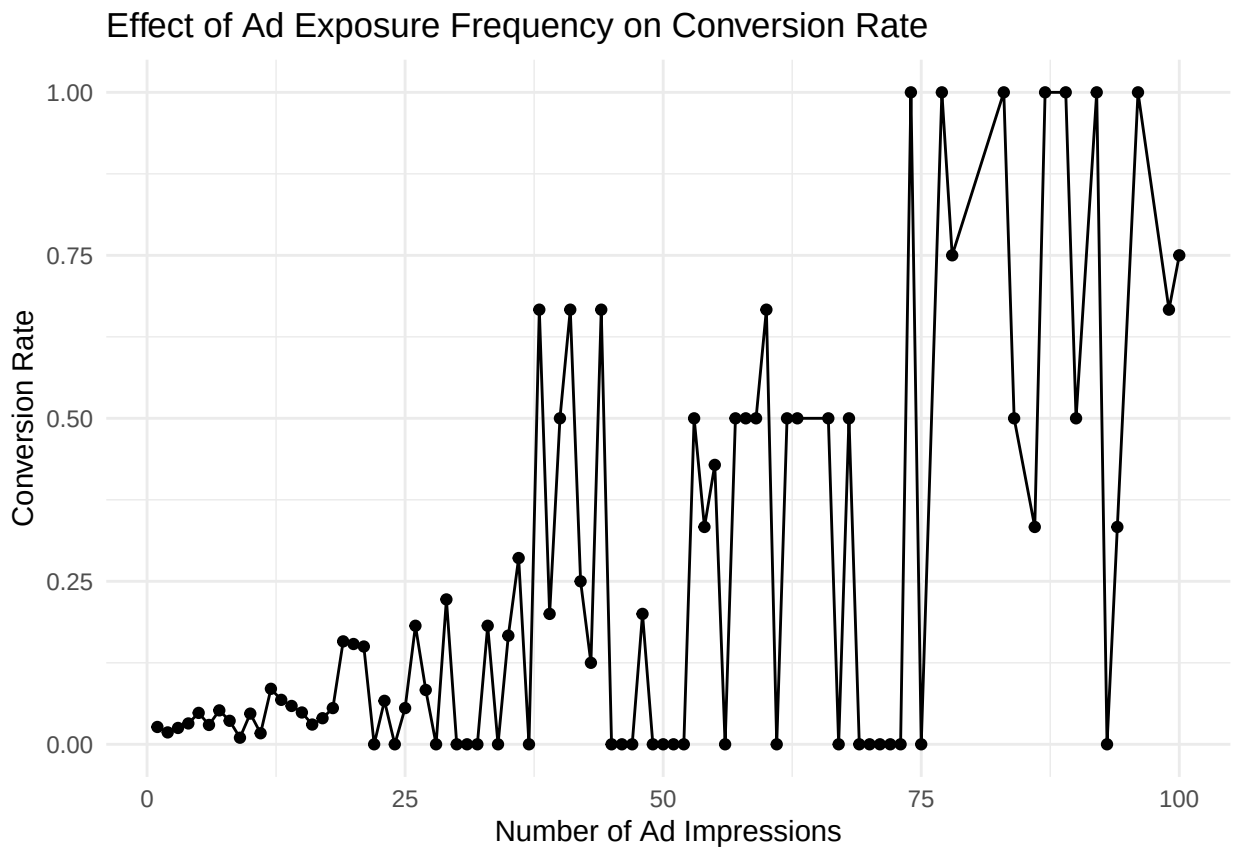
```r
net_profit
```

```
## [1] 174835.1
```

Ad Frequency Effect Analysis

```r
# Group by impressions and calculate conversion rate
impression_analysis <- df %>%
  filter(test == 1) %>%
  group_by(impressions) %>%
  summarise(conversion_rate = mean(purchase))

# Plot conversion rate vs impressions
ggplot(impression_analysis, aes(x = impressions, y = conversion_rate)) +
  geom_line() +
  geom_point() +
  labs(title = "Effect of Ad Exposure Frequency on Conversion Rate",
       x = "Number of Ad Impressions",
       y = "Conversion Rate") +
  theme_minimal()
```



Checking for Non-Compliance

```
# Check non-compliance in control group
non_compliance_control <- df %>% filter(test == 0 & impressions > 0) %>% nrow()

# Check non-compliance in treatment group
non_compliance_treatment <- df %>% filter(test == 1 & impressions == 0) %>% nrow()

# Print results
non_compliance_control
```

```
## [1] 3935
```

```
non_compliance_treatment
```

```
## [1] 0
```

Logistic Regression Model

```r
# Fit logistic regression model
logit_model <- glm(purchase ~ test + impressions, data = df, family = binomial)

# Summary of model
summary(logit_model)
```

```
##
## Call:
## glm(formula = purchase ~ test + impressions, family = binomial,
##     data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.84662    0.10397 -36.999   <2e-16 ***
## test1        0.17794    0.11267   1.579    0.114
## impressions  0.04813    0.00276  17.438   <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: 5375.8  on 19999  degrees of freedom
## Residual deviance: 5132.3  on 19997  degrees of freedom
## AIC: 5138.3
##
## Number of Fisher Scoring iterations: 6
```

Conclusion

```r
cat("The advertising campaign was profitable with a net profit of", round(net_profit, 2), "NOK.\n")
```

```
## The advertising campaign was profitable with a net profit of 174835.1 NOK.
```

```r
cat("Conversion rates increased from", round(baseline_conversion * 100, 2), "% in control to", round(co
```

```
## Conversion rates increased from 2.54 % in control to 3.1 % in treatment.
```

```r
cat("Optimal ad frequency appears to be around 5-7 impressions.\n")
```

```
## Optimal ad frequency appears to be around 5-7 impressions.
```

```r
cat("However, non-compliance in the control group may have underestimated the ad effect.\n")
```

```
## However, non-compliance in the control group may have underestimated the ad effect.
```