Assignment 3: Promotions Management

Group 31: Mayuresh Gandhi, Khyati Shah, Shafali Veeramallu, Steven Cramer, Pratik Agrawal

May 2, 2020

Contents

1 Promotional event planning

1. Is there evidence for strong seasonal demand for this product, based on the figures presented in the "Event" sheet?

Yes, the baseline numbers without promotions are high for event 1 and event 2 which were run during holiday season. This indicates strong seasonal demand during holiday period.

2. Expressed as a percentage of base sales, does event 1 or event 2 produce a greater incremental sales response?

Following are the incremental sales percentage for all events: Incremental lift for event 1: 55% Incremental lift for event 2: 69% Incremental lift for event 3: 116% Incremental lift for event 4: 14% Incremental lift for event 5: 180%

Between event 1 and 2, event 2 has higher incremental sales response.

3. Discuss the profitability results for these five events. Which events are most profitable and why? What can you learn from this analysis about what sorts of promotional events are most profitable?

After including forward buy costs in ROI, event 2 and event 5 are most profitable. Event 2, 5 and 4 are focused solely on display but event 2 and 5 have greater TPR and greater display promotions than event 4. Event 1 has higher TPR, but feature promotion is not effective, resulting in loss.

Thus, success of event 2 and event 5 can be attributed to TPR and higher display promotions. It can also be concluded that feature promotion is not very effective.

4. Recalculate the profitability for event 1 and 2 assuming that retailers will engage in 4 weeks rather than 2 weeks of forward buying.

The forward buy cost will be twice the current amount, if retailers engage in 4 weeks rather than 2 weeks of forward buying. These additional costs will change the event 1 ROI from -2% to -12%, and event 2 ROI from 44% to 34%.

5. Optional question: Compare the approach to calculating the ROI that we took in class with the Booz Allen Hamilton approach taken here (the difference was explained above). Is one of the approaches better than the other, and why?

Booz allen Hamilton's approach is understating the ROI because the incremental sales should not be counted in the cost. Without the promotions, the incremental sales would not have been obtained in first place and hence they should not be included in a variable costs. The approach used in the class is more accurate.

2 Estimating lift factors and promotion ROI analysis

```
load("Hellmans.rdata")
```

2.1 Statistics, Histograms, and Correlations

2.1.1 Creating price variable for Hellman's 32oz.mayo

```
hellmans_df$price = hellmans_df$dollars / hellmans_df$units
```

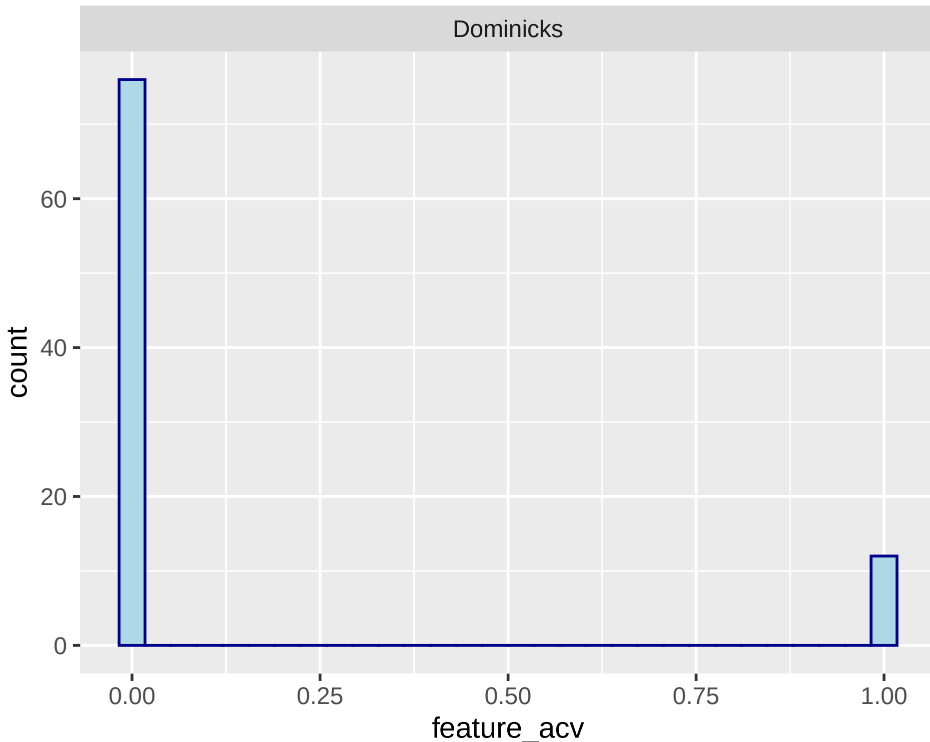
2.1.2 summarizing feature and display variables

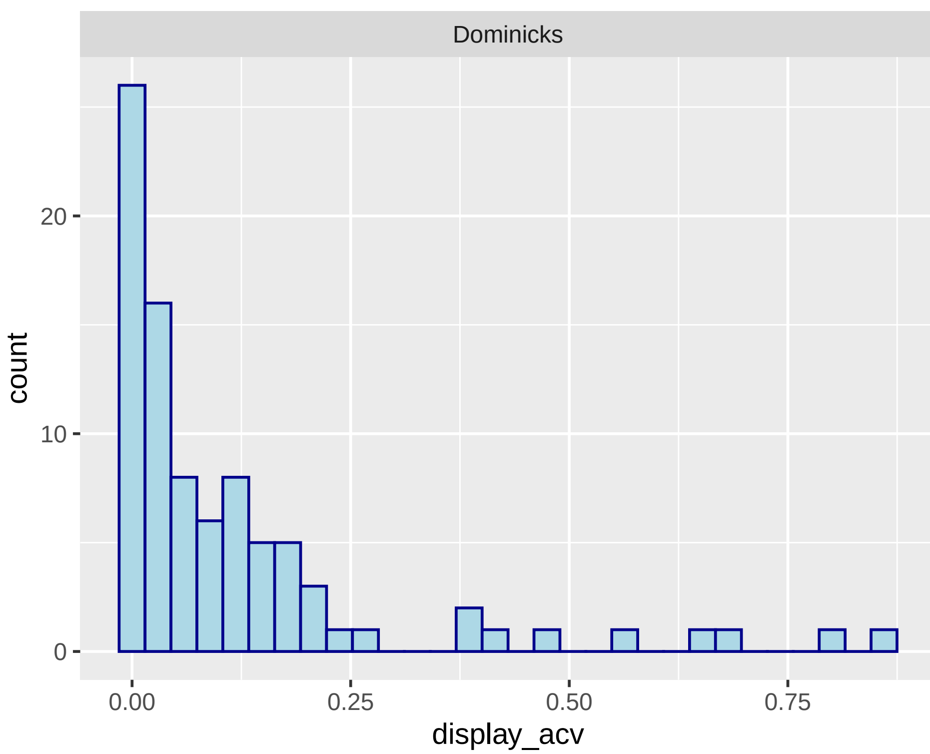
```
hellmans_df = hellmans_df %>% mutate (
                        feature_acv= feature_pctacv/100,
                        display_acv = display_pctacv/100
                       )

hellmans_df %>%
                group_by(account) %>%
                summarise(
                        number_of_observations = n(),
                        feature_mean = mean(feature_acv),
                        feature_sd = sd(feature_acv),
                        display_mean= mean(display_acv),
                        display_sd = sd(display_acv)
                       )
```

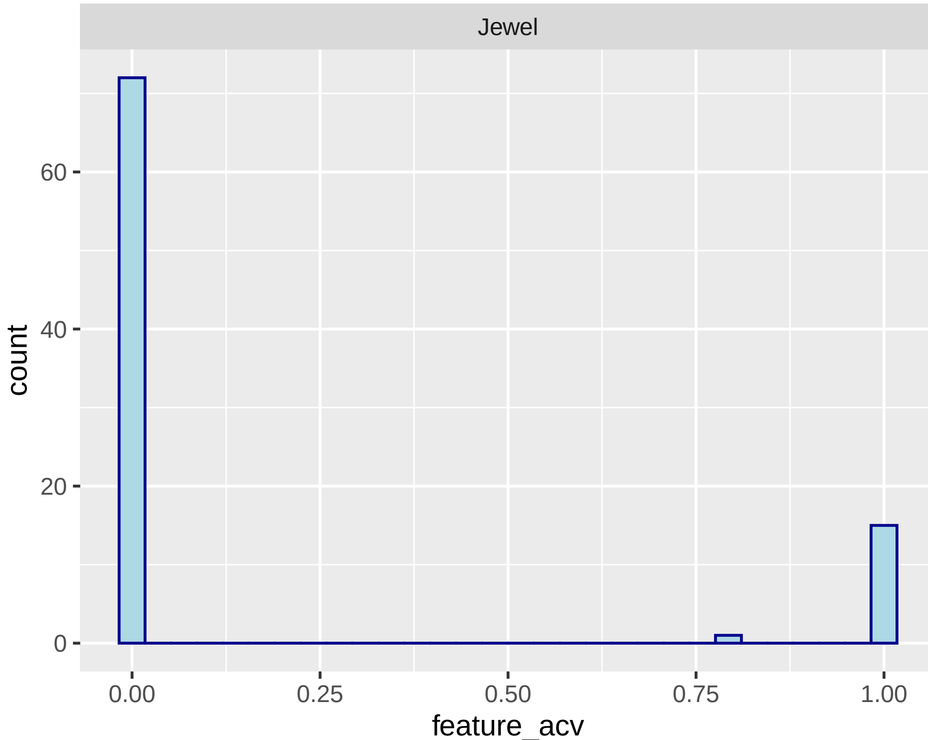
```
# A tibble: 2 x 6
  account   number_of_observati~ feature_mean feature_sd display_mean display_sd
  <chr>                    <int>        <dbl>      <dbl>        <dbl>      <dbl>
1 Dominicks                   88        0.136      0.345        0.121      0.176
2 Jewel                       88        0.179      0.383        0.230      0.258
```

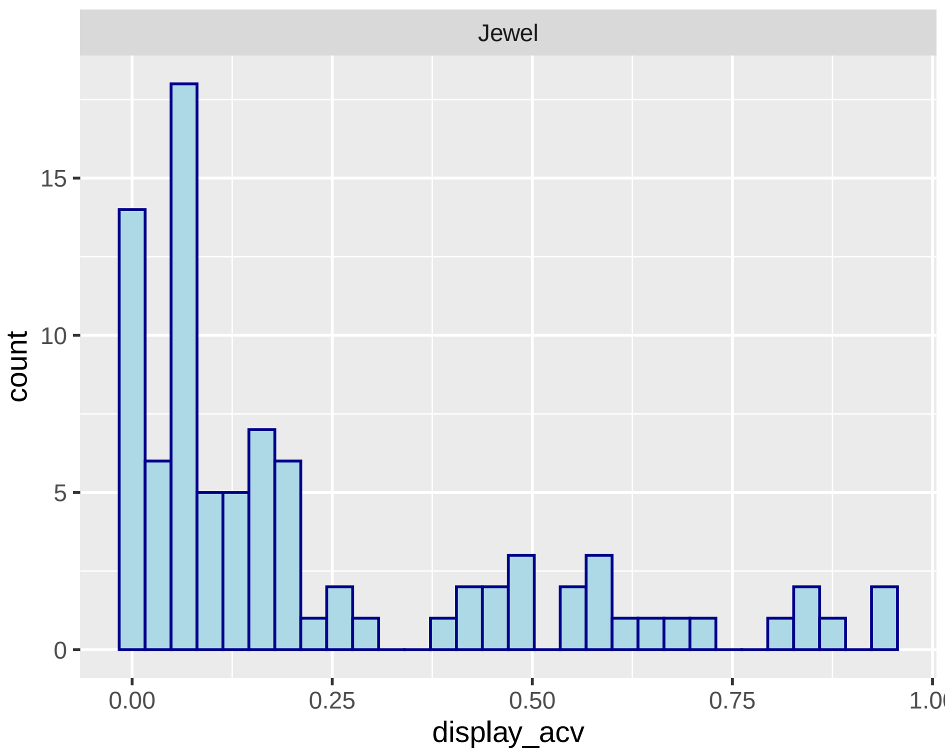
```
library(ggplot2)

for (i in unique(hellmans_df$account)) {

      plot_feature <- ggplot(hellmans_df %>% filter(account == i),
              aes(x = feature_acv)) +
              geom_histogram(color="darkblue", fill="lightblue") +
              facet_wrap(~ account)

      plot_display <- ggplot(hellmans_df %>% filter(account == i),
              aes(x = display_acv)) +
              geom_histogram(color="darkblue", fill="lightblue")+
              facet_wrap(~ account)

      print (plot_feature)
```

```
        print (plot_display)
}
```









To what extent do these two promotional instruments differ?

From the above histograms it can be inferred that there is more variation in the display attribute than in the feature based promotional attribute. Hence, we can gain more insights from display based promotion and can have more granular implementation of the promotion strategy by fine tuning the display attribute. Also, Jewel has higher promotional activities in terms of both feature and display compared to Dominick's.

2.1.3 Correlations between feature__pctacv and display__pctacv

```
hellmans_df %>% select(feature_acv,display_acv,price) %>% cor()
```

```
            feature_acv display_acv      price
feature_acv   1.0000000   0.7599992 -0.5747241
display_acv   0.7599992   1.0000000 -0.6700056
price        -0.5747241  -0.6700056  1.0000000
```

Comment on your findings.

Feature and display are highly correlated. There is high inverse correlation between feature-display and price too. Because feature and display are highly correlated, it is difficult to ascertain whether the correlation of price is due to one or both of these influencing terms.

Do the correlations indicate a potential problem for your regression analysis to be performed below?

Yes, they do. These highly correlated attributes will introduce bias in the estimates unless they are accounted for.

2.2 Log linear demand model

2.2.1 Demand Model Using only price for each account

```
for (i in unique(hellmans_df$account))
{
    m = hellmans_df %>% filter (account == i) %>% lm(log (units) ~ log (price), .)
    print(paste("Linear log model for : ", i, sep=""))
    print (summary(m))
}
```

```
[1] "Linear log model for : Dominicks"

Call:
lm(formula = log(units) ~ log(price), data = .)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.63760 -0.17214 -0.01628  0.10558  0.79349 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  10.0368     0.0719  139.60  < 2e-16 ***
log(price)   -4.1665     0.4107  -10.15  2.3e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2572 on 86 degrees of freedom
Multiple R-squared:  0.5448,	Adjusted R-squared:  0.5395 
F-statistic: 102.9 on 1 and 86 DF,  p-value: 2.304e-16

[1] "Linear log model for : Jewel"

Call:
lm(formula = log(units) ~ log(price), data = .)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.59230 -0.16883 -0.03486  0.15152  0.81131 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.60443    0.05259  201.66   <2e-16 ***
log(price)  -4.58359    0.42660  -10.74   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2482 on 86 degrees of freedom
Multiple R-squared:  0.5731,	Adjusted R-squared:  0.5681 
F-statistic: 115.4 on 1 and 86 DF,  p-value: < 2.2e-16
```

2.2.2 Demand Model Adding feature and display variables in addition to price

```r
# Using the original feature_pctacv and display_pctacv variables:

for (i in unique(hellmans_df$account))
{
    m = hellmans_df %>% filter (account == i) %>% lm(log (units) ~ log (price) + feature_pctacv + displa
    print(paste("Linear log model for : ", i, sep=""))
    print (summary(m))
}
```

```
[1] "Linear log model for : Dominicks"

Call:
lm(formula = log(units) ~ log(price) + feature_pctacv + display_pctacv, 
    data = .)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.34144 -0.13771 -0.02137  0.11067  0.61078 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     9.5212338  0.0894423 106.451  < 2e-16 ***
log(price)     -1.8431810  0.4503167  -4.093 9.74e-05 ***
feature_pctacv  0.0028531  0.0008925   3.197  0.00196 ** 
display_pctacv  0.0083410  0.0017653   4.725 9.14e-06 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2021 on 84 degrees of freedom
Multiple R-squared:  0.7255,    Adjusted R-squared:  0.7157
F-statistic:    74 on 3 and 84 DF,  p-value: < 2.2e-16

[1] "Linear log model for : Jewel"

Call:
lm(formula = log(units) ~ log(price) + feature_pctacv + display_pctacv,
    data = .)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36769 -0.12020 -0.02219  0.08526  0.49093

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.0888107  0.0632725 159.450  < 2e-16 ***
log(price)     -1.8973532  0.3996928  -4.747 8.39e-06 ***
feature_pctacv -0.0009124  0.0009062  -1.007    0.317
display_pctacv  0.0106947  0.0014891   7.182 2.56e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1702 on 84 degrees of freedom
Multiple R-squared:  0.8039,    Adjusted R-squared:  0.7969
F-statistic: 114.8 on 3 and 84 DF,  p-value: < 2.2e-16
```

Comment on the difference between the two regressions in terms of goodness of fit, and the price elasticity estimates. Is the change in price elasticity estimates as expected? What is the reason for this change? Are the coefficient estimates similar for both accounts?

The F-statistics is significant for all the 4 regressions. Hence, the overall regressions are statistically significant, indicating that R-squared can be used to assess goodness of fit. R-squared increases when feature and display are added as predictor variables to 72% (Dominicks) and 80% (Jewel), indicating improvement in goodness of fit.

The price elasticities without Feature and display are high which can indicate that there are more predictors that may explain the change in demand. Evidently, adding feature and display reduces price elasticies at both Jewel and Dominick. This is due to the fact that estimates from Features and display removes the promotion bias from price elasticity.

With modified regressions, price elasticities at both accounts are very similar indicating there is no impact on Hellman's Mayo due to stores. However, we can see that coefficient of features is more significant at Dominicks (lower standard error). Features are insignificant at Jewel where as displays are slightly more significant at Jewel than at Dominicks.

2.3 Lift Factors for three promotions

Calculate the lift factors for each promotion for both accounts, based on the regression estimates in 2. Set estimates that are not statistically significant = 0.

2.3.1 Promotion 1: 15% TPR

```r
promo_1_price = -0.15
promo_1_feature = 0
promo_1_display = 0

# Regression Coefficients for Dominick's:
dm_price_elasticity = -1.843181
dm_feature_coeff = 0.0028531
dm_display_coeff = 0.008341

# Lift Factor for Dominick's

dm_lift_factor_1 = exp(
  dm_price_elasticity * log (1 + promo_1_price) +
  dm_feature_coeff * promo_1_feature +
  dm_display_coeff * promo_1_display
)

print(paste("Lift Factor for Dominick's from Promo 1 is: ", dm_lift_factor_1, sep=""))
```

```
[1] "Lift Factor for Dominick's from Promo 1 is: 1.34925394045533"
```

```r
# Regression Coefficients for Jewel:
j_price_elasticity = -1.8973532
j_feature_coeff = 0
j_display_coeff = 0.0106947

# Lift Factor for Jewel

j_lift_factor_1 = exp(
  j_price_elasticity * log (1 + promo_1_price) +
  j_feature_coeff * promo_1_feature +
  j_display_coeff * promo_1_display
)

print(paste("Lift Factor for Jewel from Promo 1 is: ", j_lift_factor_1, sep=""))
```

```
[1] "Lift Factor for Jewel from Promo 1 is: 1.36118522738297"
```

2.3.2 Promotion 2: 15% TPR, 70% display

```r
promo_2_price = -0.15
promo_2_feature = 0
```

```r
promo_2_display = 70

# Regression Coefficients for Dominick's:
dm_price_elasticity = -1.843181
dm_feature_coeff = 0.0028531
dm_display_coeff = 0.008341

# Lift Factor for Dominick's

dm_lift_factor_2 = exp(
  dm_price_elasticity * log (1 + promo_2_price) +
  dm_feature_coeff * promo_2_feature +
  dm_display_coeff * promo_2_display
)

print(paste("Lift Factor for Dominick's from Promo 2 is: ", dm_lift_factor_2, sep=""))
```

```
[1] "Lift Factor for Dominick's from Promo 2 is: 2.41916346064906"
```

```r
# Regression Coefficients for Jewel:
j_price_elasticity = -1.8973532
j_feature_coeff = 0
j_display_coeff = 0.0106947

# Lift Factor for Jewel

j_lift_factor_2 = exp(
  j_price_elasticity * log (1 + promo_2_price) +
  j_feature_coeff * promo_2_feature +
  j_display_coeff * promo_2_display
)

print(paste("Lift Factor for Jewel from Promo 2 is: ", j_lift_factor_2, sep=""))
```

```
[1] "Lift Factor for Jewel from Promo 2 is: 2.87768114239633"
```

2.3.3 Promotion 3: 15% TPR, 70% display, 100% feature

```r
promo_3_price = -0.15
promo_3_feature = 100
promo_3_display = 70

# Regression Coefficients for Dominick's:
dm_price_elasticity = -1.843181
dm_feature_coeff = 0.0028531
dm_display_coeff = 0.008341

# Lift Factor for Dominick's

dm_lift_factor_3 = exp(
  dm_price_elasticity * log (1 + promo_3_price) +
```

```r
  dm_feature_coeff * promo_3_feature +
  dm_display_coeff * promo_3_display
)

print(paste("Lift Factor for Dominick's from Promo 3 is: ", dm_lift_factor_3, sep=""))
```

```
[1] "Lift Factor for Dominick's from Promo 3 is: 3.21790910700867"
```

```r
# Regression Coefficients for Jewel:
j_price_elasticity = -1.8973532
j_feature_coeff = 0
j_display_coeff = 0.0106947

# Lift Factor for Jewel

j_lift_factor_3 = exp(
  j_price_elasticity * log (1 + promo_3_price) +
  j_feature_coeff * promo_3_feature +
  j_display_coeff * promo_3_display
)

print(paste("Lift Factor for Jewel from Promo 3 is: ", j_lift_factor_3, sep=""))
```

```
[1] "Lift Factor for Jewel from Promo 3 is: 2.87768114239633"
```

2.4 ROI Analysis

Please refer to the ROI Analysis excel sheet for detailed ROI calculations.