Marketing Models Project

Report by Group 10 for Pricing Project: 04/19/2019

- Chengzhang Jiang
- Eric Greenberg Goldy
- Haoyang Sang
- Pooja Adiseshu

1. Introduction

Marketers need to align the correct channels through which to promote their product in order to sell in the most efficient way possible to drive revenue and profit. In order to do so, it is imperative to collect an understanding, through analyzing data on the marketing "vehicle," of how to best promote the product. There are many vehicles through which promotion can take place. Most traditionally this includes TV and Radio advertisements. These forms of promotion were difficult to completely assess due to the ambiguous and indirect relation to sales. With the growth of the web though this problem is easier to assess. There is available data to analyze web tracking, click-throughs, and measures on whether a consumer watched a video ad in its entirety.

This type of information is invaluable to an organization, as they can use it to better penetrate their customers. With detailed data on the success of various ad types, organizations learn more about which types of ads work for certain customers. Organization can segment their customers on a more individual level, and make use of target promotions, which increase the efficiency of a sale. This report aims to analyze several different marketing vehicles and attempt to describe their various effects on sales for Mahou San Miguel.

2. Background

Mahou San Miguel, a Spanish brewer, sells three San Miguel beer products in Mahou San Miguel, a Spanish brewer, sells three San Miguel beer products in Pernalonga stores and regularly partners with Pernalonga to promote its products via weekly flyers and in-store displays. Mahou San Miguel also employs other marketing vehicles such as email, web (display banners and paid search), and traditional media (TV and Radio).

Mahou San Miguel wants to verify the effectiveness of its promotions and marketing partnership with Pernalonga. He is interested in identifying promotion and marketing activities that drive significant incremental sales for continuation into 2018.

The approach we took to solve this problem is detailed in the report. We are using different statistical methods to analyze the historical data, to estimate the impact of various marketing vehicles on the sales volume of Beer. To do this, we followed the steps

3. Data and Model Preparation

Section 1: In order to prepare the data for this analysis, we had to walk through several steps. The steps to prepare the data for our code go as the following.

1) Firstly, we identified the three products Mahou San Miguel sells. We then computed the Reach for both TV and Radio for the three products.
2) Noticing that TV and Radio promotions are targeted towards all products we calculated the GRP and impressions for each week

Section 2: Used the half life provided to adstock the GRP with missing weeks filled in and then transformed the adstock GRP to reach to run the regression

1) We calculated the decay parameter of TV commercial based on TV half-life (8 weeks)

$$\alpha = 1 - (0.5) \wedge (1/8)$$

2) Used this decay parameter to convert GRPs into AdStock GRPs

   *AdStock GRPs at week t = αGRPs in week t + (1 – α) * AdStock GRPs at week t-1*

3) Converted adstocked GRP to reach (2+ Reach)

$$\% \text{ Reach} = 0.90(1\text{-}e \wedge (-0.025 * \text{AdStock GRPs}))$$

4) Calculate GRP and impression of radio each week
5) Calculate alpha of radio commercial based on radio half-life (4 weeks)

$$\alpha = 1 - (0.5) \wedge (1/4)$$

6) Used this decay parameter, alpha to convert GRPs into AdStock GRPs

   *AdStock GRPs at week t = αGRPs in week t + (1 – α) * AdStock GRPs at week t-1*

7) Convert adstocked GRP to reach (2+ Reach)

$$\% \ Reach = 0.95(1-e \wedge (-0.020 * AdStock \ GRPs))$$

Section 3: Prepare the weekly data using the transaction, seasonality, and holiday tables. In terms of holidays, it was split into two columns holiday and important holiday, which was for Christmas and New Years. When plotting the sales across weeks, we noticed there are patterns at the turning point of each year that are very different from other holidays. Looking at product 138936952, it's sales in this period plummeted for some reason.

1) Determined list price and discount
2) Merged transaction table with seasonality and holiday table
3) Graphed product

Section 4: Fit the model for the three products

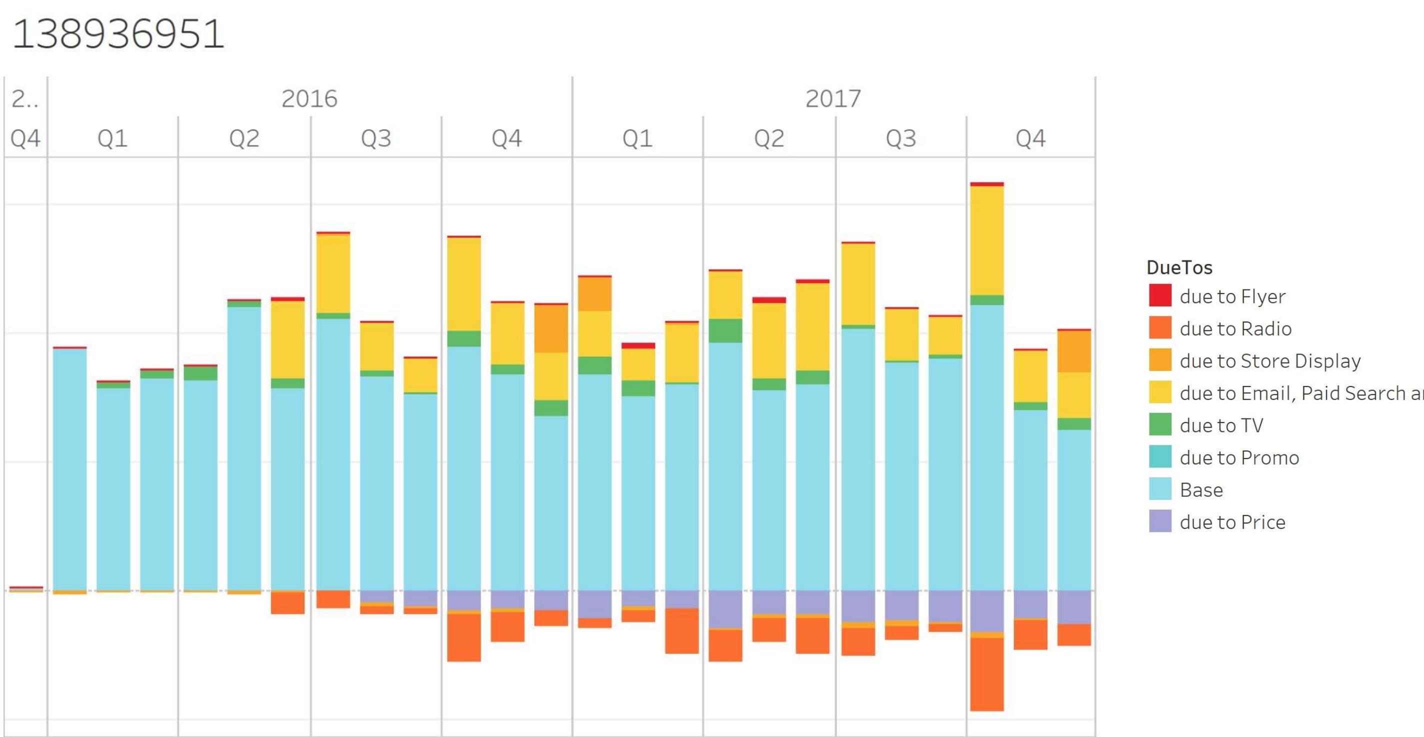
1) We merged promotions with TV and Radio
2) Looked at uniqueness of the different vehicles
3) Used the earliest price as the base price
4) Transformed the sales quantity variable using the assumption that the maximum sales quantity is 10% more than the historical maximum
5) Run regressions for each product with transformed variables
6) Merge the transaction information of a specific product with all the promotion advertisements associated
7) Apply a logit function (because it is bounded and will perform reasonably within range) which is ideal for our situation and helps depict complex implicit interaction, which is very likely occurring when the audience is exposed to a lot of similar media such as TV and Radio

Section 5: Calculate Dueto:

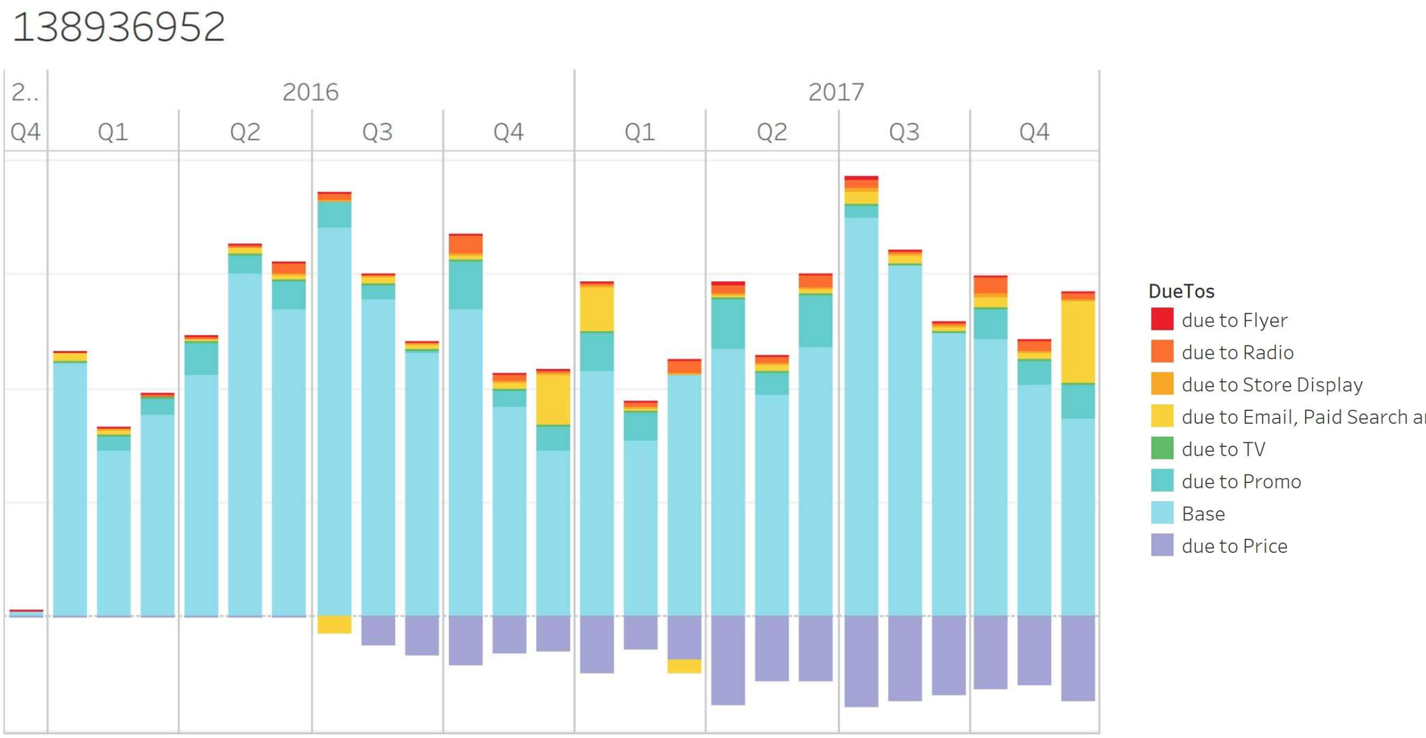
1) Created the baseline as the base price plus seasonality, and holiday. Used the price at the beginning of the two year period as the base price with other variables being replaced with 0

2) Dueto price here is the difference between the quantities estimated using the real price in that period and the quantities estimated with the base price in place of the weekly price while the value of the other variables remain the same
3) The process follows similar to the above, except we must rescale the data back to its original quantity
4) We then conduct a Durbin Watson test to calculate autocorrelation
5) Repeat for the remaining products

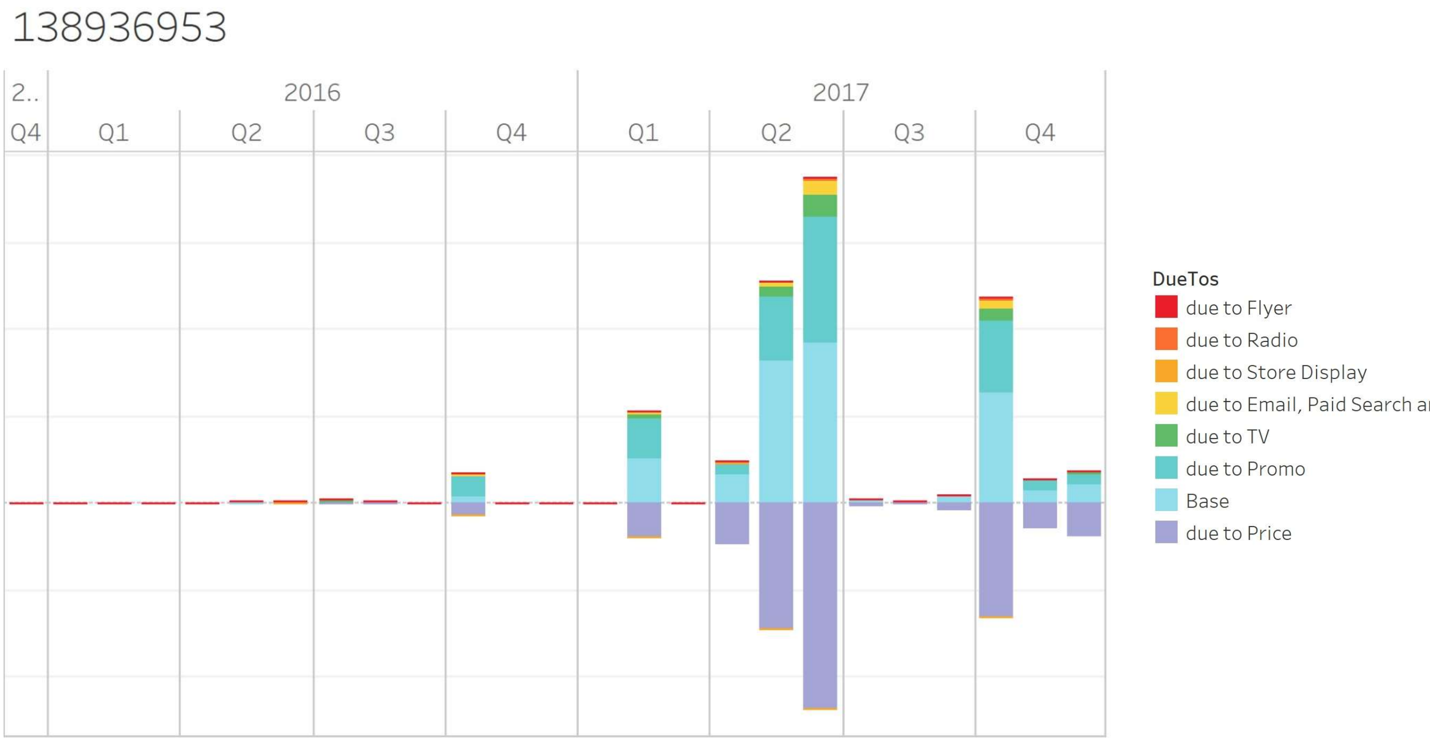
4. Results



The plot above depicts shows the stacked bar chart for the various DueTos in 2016-2017. We see that Email, Paid Search, and Web Display make up a large portion of the positive effects. On the other hand radio seems to be negatively affecting sales for this product.



The plot above shows a stacked bar chart for product 138936952. It is notable that the base makes up a substantially higher portion than compared to the other products. Email Paid Search and Web Display also make up the largest portion of positive product performance as well here, and promos seem to have a strong positive affect.



Promotion is nearly the same size as the base for this product, indicating that it relies heavily on promotion. It also appears to be only sold at particular times, so is a very seasonal product.

5. Statistical Model Diagnostics

The formula we are fitting on is:

sales(transformed) ~ wkly_price + wkly_dct + Flyer + Email + Web_Display + Paid_Search + TV + Radio + seas_index + holiday + imt_hol

For item 138926951, 138926952 and 138926953, we run the linear regression on the transformed logit functions and the result is shown below. It shows that all three model have significant F values, meaning that the factors combined have significant impact on changing the sale.

For product 138926951, the model shows that weekly price, Email, TV commercial and important holidays (Christmas and New Year) has significant on sale with p value less than 0.05;

For product 138926952, the model shows that weekly price, weekly discount, Email, seasonality effect and important holidays (Christmas and New Year) have significant on sale with p value less than 0.05;

For product 138926953, the model shows that weekly price, weekly discount, store display and seasonality effect have significant on sale with p value less than 0.05.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.435e+00  1.257e+00   1.938 0.055683 .  
wkly_price  -3.537e+00  1.717e+00  -2.061 0.042104 *  
wkly_dct     3.563e+00  1.997e+00   1.784 0.077615 .  
Flyer        3.140e-01  4.038e-01   0.778 0.438740    
Email        6.304e-06  1.780e-06   3.541 0.000622 ***
Web_Display  3.314e-07  6.879e-07   0.482 0.631117    
Paid_Search -1.512e-06  4.072e-06  -0.371 0.711216    
TV           2.172e+00  9.950e-01   2.183 0.031513 *  
Radio       -1.046e+00  6.945e-01  -1.506 0.135459    
seas_index   3.906e-05  3.919e-05   0.997 0.321524    
holiday1    -9.925e-02  1.780e-01  -0.558 0.578478    
imt_hol     -2.451e+00  4.273e-01  -5.737 1.17e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5506 on 94 degrees of freedom
Multiple R-squared:  0.4748,	Adjusted R-squared:  0.4133 
F-statistic: 7.725 on 11 and 94 DF,  p-value: 2.303e-09
```

result for 51

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    2.475e+00  9.970e-01   2.482 0.014849 *  
wkly_price    -1.011e+00  2.404e-01  -4.205 6.00e-05 ***
wkly_dct       1.053e+00  2.915e-01   3.615 0.000488 ***
Flyer          4.313e-02  1.642e-01   0.263 0.793352    
Email          4.693e-06  1.452e-06   3.233 0.001698 ** 
Web_Display   -6.148e-07  5.617e-07  -1.095 0.276544    
Paid_Search    2.099e-06  3.268e-06   0.642 0.522240    
Store_Display  1.179e-01  2.362e-01   0.499 0.618934    
TV             2.531e-02  8.024e-01   0.032 0.974901    
Radio          2.105e-01  5.581e-01   0.377 0.706948    
seas_index     1.774e-04  3.222e-05   5.506 3.23e-07 ***
holiday1       2.287e-02  1.385e-01   0.165 0.869138    
imt_hol       -1.397e+00  3.421e-01  -4.083 9.41e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4448 on 93 degrees of freedom
Multiple R-squared:  0.6271,	Adjusted R-squared:  0.579 
F-statistic: 13.04 on 12 and 93 DF,  p-value: 3.05e-15
```

result for 52

Both product 51 and 52's sale are highly dependent on important holidays, and Email campaigns, while both 52 and 53 are season-sensitive products, therefore we can tell that people buy them for different reasons and they belong to different product lines.

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.5154 -0.4196  0.1181  0.4310  1.1277

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.806e+00  1.649e+00   4.734 8.27e-06 ***
wkly_price    -8.516e-01  1.068e-01  -7.977 4.85e-12 ***
wkly_dct       7.093e-01  1.555e-01   4.561 1.62e-05 ***
Flyer          2.112e-03  3.168e-01   0.007   0.9947
Email          1.613e-07  2.039e-06   0.079   0.9371
Web_Display    5.021e-07  8.085e-07   0.621   0.5362
Paid_Search    4.834e-06  4.803e-06   1.006   0.3170
Store_Display -1.421e+00  6.381e-01  -2.227   0.0285 *
TV             4.574e-01  1.170e+00   0.391   0.6968
Radio          6.891e-02  8.212e-01   0.084   0.9333
seas_index     3.755e-04  4.633e-05   8.104 2.66e-12 ***
holiday1      -1.310e-01  2.057e-01  -0.637   0.5259
imt_hol       -4.332e-01  5.034e-01  -0.861   0.3918
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6316 on 89 degrees of freedom
Multiple R-squared:  0.7752,    Adjusted R-squared:  0.7449
F-statistic: 25.58 on 12 and 89 DF,  p-value: < 2.2e-16
```

result for 53

Another interesting finding is that all three products' sale are positively affected by weekly discount and negatively affected by weekly price in a significant fashion. It shows that the customers not only are sensitive to the pocket price but also the list price of the product. An explanation is that people grow accustomed to discounts so that more people are waiting for the discount: when price goes back to normal they choose not to buy any of them.

Also, to test the autocorrelation we used the Durbin Watson statistic test. is a test for autocorrelation in the residuals from a statistical regression analysis. The Durbin-Watson statistic will always have a value between 0 and 4. A value of 2.0 means that there is no autocorrelation detected in the sample. Values from 0 to less than 2 indicate positive autocorrelation and values from from 2 to 4 indicate negative autocorrelation. All three products' time series show a DW value of almost 2, which means there is very likely no autocorrelation within the observation.

```
        Durbin-Watson test

data:  model$sales_trfm ~ model$residual
DW = 1.995, p-value = 0.4725
alternative hypothesis: true autocorrelation is greater than 0
```

DW test for 138936951: 1.995

```
        Durbin-Watson test

data:  model$sales_trfm ~ model$residual
DW = 1.9406, p-value = 0.3602
alternative hypothesis: true autocorrelation is greater than 0
```

DW test for 138936952: 1.9406

```
        Durbin-Watson test

data:  model$sales_trfm ~ model$residual
DW = 1.9046, p-value = 0.2892
alternative hypothesis: true autocorrelation is greater than 0
```

DW test for 138936953: 1.9046

6. Future Improvements

Future Interactive DueTo calculation

Since logit models have complex interactions between the causal values, when we explain the logit model we are also expected to explain complex implicit interactions.

There is one heuristic way to calculate the interaction effect of A and B: We can calculate the DueTo of A, B and $A \cup B$ by removing them from the model accordingly. Then we can calculate the interaction $A \cap B = A + B - A \cup B$

However due to the various number of features it is neither feasible nor sensible to calculate the DueTos of all interactions larger than 3. The interaction effect of A, B and C will be $A \cap B \cap C = A + B + C + A \cap B + A \cap C + B \cap C - A \cup B \cup C$. Neither can we to explain the interaction in such an easy way.