Marketing Mix Models

Team 6

Joyce Zhang, Nicole Santolalla, Justin Wood

April 16th, 2021

Table of Contents

Introduction

Background

Pernalonga, a leading supermarket chain of over 400 stores in Lunitunia, sells over 10 thousand products in over 400 categories. Pernalonga regularly partners with suppliers to fund promotions and derives about 30% of its sales on promotions. Until recent experimentation with personalized promotions, most of Pernalonga's promotions are chain-wide promotions. The beer category is promoted regularly in Pernalonga's stores. Mahou San Miguel, a Spanish brewer, sells three San Miguel beer products in Pernalonga's 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).

Proposed approach

Our goal is to verify the effectiveness of its promotions and marketing partnership with Pernalonga. The client is interested in identifying promotion and marketing activities that drive significant incremental sales for continuation into 2021 for Mahou San Miguel. We will focus on 3 Mahou San Miguel beer products that has been sold in Pernalonga in 2016 and 2017. We will explore marketing mix modeling in the Lunitunia market to find the effect of seasonality of the beer category, national holidays in Lunitunia in 2016 and 2017, and other vehicles on the products' sales quantity. Then we will recommend which promotion on each product for Mahou San Miguel and Pernalonga to continute in 2021.

Methodology

Business Context

The Beer Segment accounted for 40% of the Alcoholic Drinks market revenue in 2019. Revenue in 2019 worldwide was 650.7 billion worldwide and expected to grow 867.6 billion (33%) in revenue by 2025. Key players in the Alcoholic Drinks Market include Anheuser-Busch, Heineken, Asahi, San Miguel Corp, and Top Frontier Investment Holdings. In 2019, San Miguel Corp was ranked 4 as having the highest total revenue of the top companies in the Alcoholic Drinks market, which includes all types of alcohol[1]. To get to San Miguel's beer ranking and insights, we need to dive and explore into the data. San Miguel is a 100% spanish owned family company with production share of 32%. The company has 10 different brewery centers, 8 in Spain and two in the USA. San Miguel produce 70% of the spanish beer consumed outside of Spain, which makes San Miguel the most international beverage company in Spain[2].

Data Understanding

We explored the transactions of 3 San Miguel beer products sold in Pernalonga, seasonality, holiday, and promotion activities of the 3 products. This analysis is based on **weekly data**, starting from 2016-01-03 till 2017-12-31. First, we computed discount rate as total discount amount divided by total sales amount. We looked at the 3 products individually and compute the following statistics.

Product ID	Total Sales Quantity	Total Sales	Mean Unit Price	Mean Discount Rate
138936951	22,477	$16,767.35	$0.746	4.5%
138936952	10,895	$46,670.55	$4.28	4.92%
138936953	1,261	$19,707.39	$15.63	7.89%

Figure 1. Summary Statistics of 3 San Miguel products sold in Pernalonga 2016 and 2017

Additionally, in 2016 and 2017, product 138936951 was bought by around 3.3k customers and sold in 406 stores; product 138936952 was bought by around 2.8k customers and sold in 400 stores; product 138936953 was bought by around 595 customers and sold in 283 stores. Based on the facts of each product, we assumed that product 138936951 is the most popular product that sold the most with least discount rate, while product 138936953 is the opposite.

[1] Source: file:///C:/Users/james_/Downloads/study_id48816_beer-report.pdf
[2] source:https://www.mahou-sanmiguel.com/en-gb/about-us

The promotion activity data has information for seven different types of marketing vehicles, but only five of them have been promoted on all 3 products: TV, Radio, Paid Search, Web Display and Email. Flyer and Store Display were only promoted on certain products. The effectiveness of each marketing vehicles will be quantified differently. We will use GRP (Gross Rating Points) to measure TV and Radio promotions

Data Preparation

First, we looked seasonality to check if it covers the breadth of both 2016 and 2017 on weekly basis. The holiday seasonality index started from week of December 27th, 2015, while the transaction data started from January 2nd, 2016. We performed transformation of the transaction date on weekly basis to match the transaction week in seasonality table. Then, we created two variables *weekly shelf price* and *weekly discount rate*.

- **Weekly Shelf Price**: we calculated the shelf price of each product on a weekly basis, the shelf price is the weighted average list price.
- **Weekly Discount Rate**: we calculated the discount rate of each product on a weekly basis, the discount rate of the product is the total discount amount / total sales amount.

Before building the marketing mix models, we need to transform the quantity of promotion effectiveness of Radio and TV. We used reach as the measurement as it is robust, and commonly used in marketing practice. First we need to calculate the decay parameter, alpha for both TV and Radio. We know that TV advertisement have an 8-week half-life and Radio ads have 4-week half-life. Then we computed the AdStock GRPs for the period impacted by TV and Radio with the following formula.

$$AdStock_t = \alpha GRP_t + (1-\alpha)AdStock_{t-1}$$

Then we converted AdStock GRPs to Reach on a weekly level with the following formulas to use Reach as input in our marketing model.

$$Reach_{TV} = 0.95(1 - e^{-0.020(AdStock\ GRPs)})$$

$$Reach_{Radio} = 0.90(1 - e^{-0.025(AdStock\ GRPs)})$$

We have other five marketing vehicles: Paid Search, Web Display and Email, Flyer and Store Display. Flyer and Store Display amount in the promotion activity dataset was set to either 0 or 1, we set them as binary variable, with 1 indicating that this marketing vehicle was used that week, 0 otherwise. For the other vehicles, we simply used amount as measurement input.

Finally, we merged all the marketing vehicles input with weekly shelf price, weekly discount rate, holiday and seasonality index and built marketing mix models for each product.

Modeling & Evaluation

Modeling Approach

For modeling, we decided to use logit model to capture the complexity of the interactions between independent variables. The dependent variable, weekly quantity sold is bounded and we took a limit of 10% of the maximum volume as maximum limit. This model needs heuristics to calculate the sale value, but it performs better at modeling the relation among the variables.

Modeling Results Evaluation

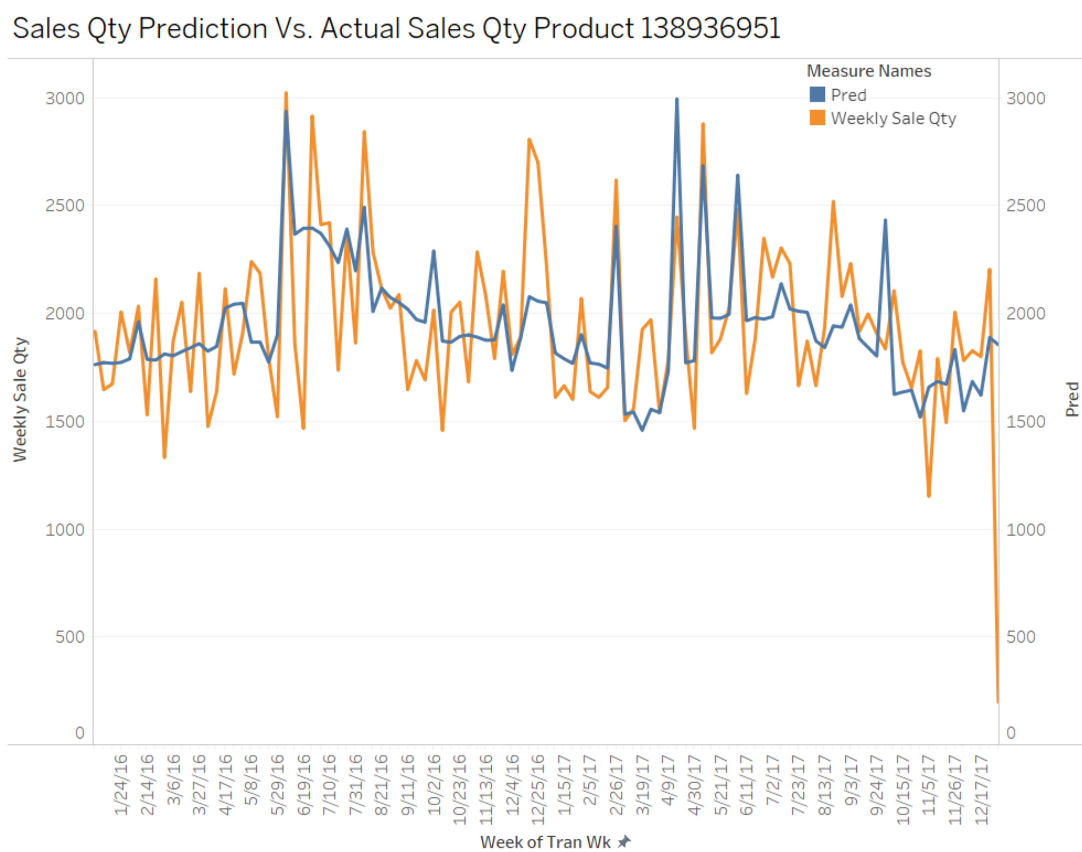
After apply logit transformation on the weekly sales quantity and standardize/normalize the input marketing vehicles and other features in the model, we generated the following evaluation table calculated from the predicted weekly sales quantity and the actual weekly sales quantity.

Product ID	R Squared	MAE	MAPE	RMSE
138936951	0.36	26.58	0.17	36.03
138936952	0.69	13.44	0.13	17.49
138936953	0.79	3.55	0.45	4.94

Figure2. Models evaluation for 3 products

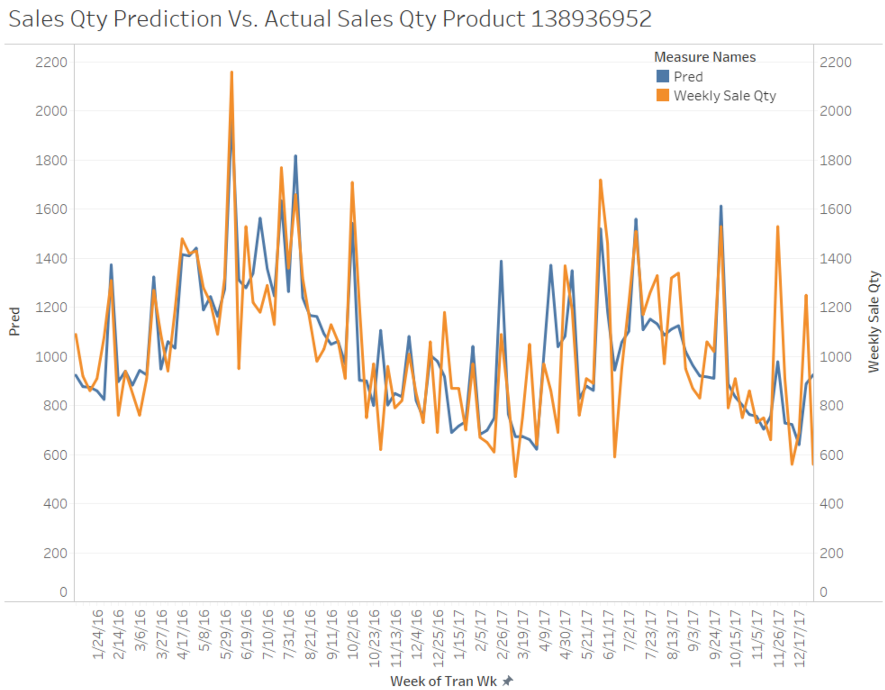
Overall, the models performed the best with model on product 138936953, with relatively high r squared and lower RMSE and MAE. Then we will look at each product individually to interpret the coefficient of the model.

Product 138936951



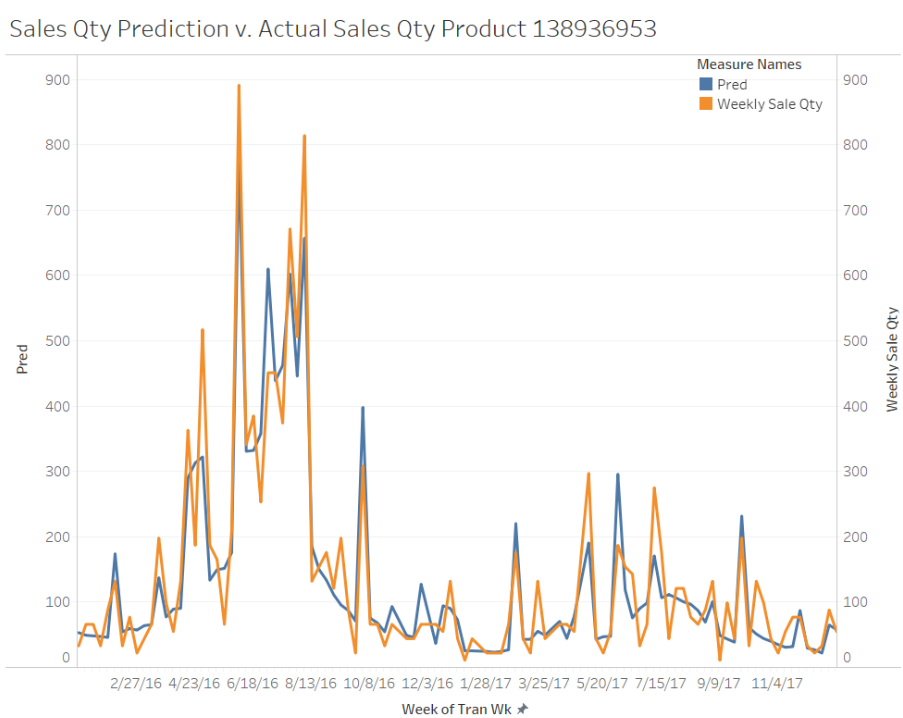
From the model results, intercept is the most significant feature as expected, since base forms a large part of the sales quantity. Weekly shelf price is the next most significant variable which has a negative effect on the quantity. TV is the next significant variable which has a positive effect on the quantity.

Product 138936952



From the model results, weekly shelf price, weekly discount rate and seasonality index are the most important features. Weekly shelf price has a negative impact on weekly sales quantity, while the weekly discount rate and seasonality index have positive impact.

Product 138936953



From the model results, weekly shelf price, weekly discount rate, seasonality index are significant features. Weekly shelf price has negative impact on weekly sales quantity, while weekly discount rate and seasonality index have positive impact on weekly sales quantity.

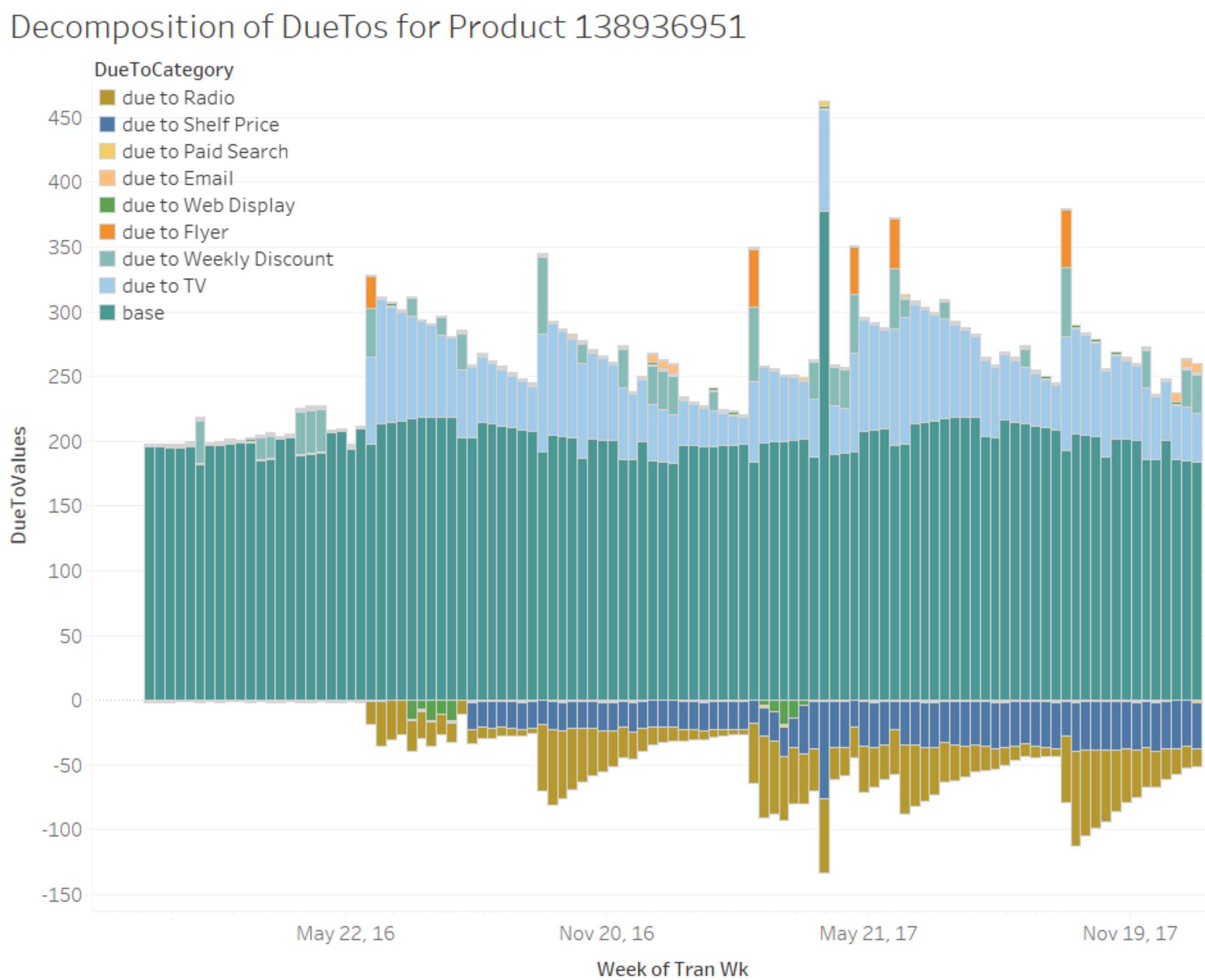
Decomposition of DueTos

After we finished modeling for each product, we computed the DueTos to decompose weekly sales quantity into the base sales quantity that Pernalonga can sell without any marketing activities. First, we needed to set a base for sales quantity. We defined our base sales quantity for each product as the base value adding up with seasonality and holiday, since seasonality and holiday are not factors that we can normally avoid. The base sales quantity was computed by setting all the other marketing vehicles to 0, then run the logit regression and log transform the result. Then we needed to figure out the base value of each marketing vehicles. We chose to set the base price of the product as the first shelf price from transactions.

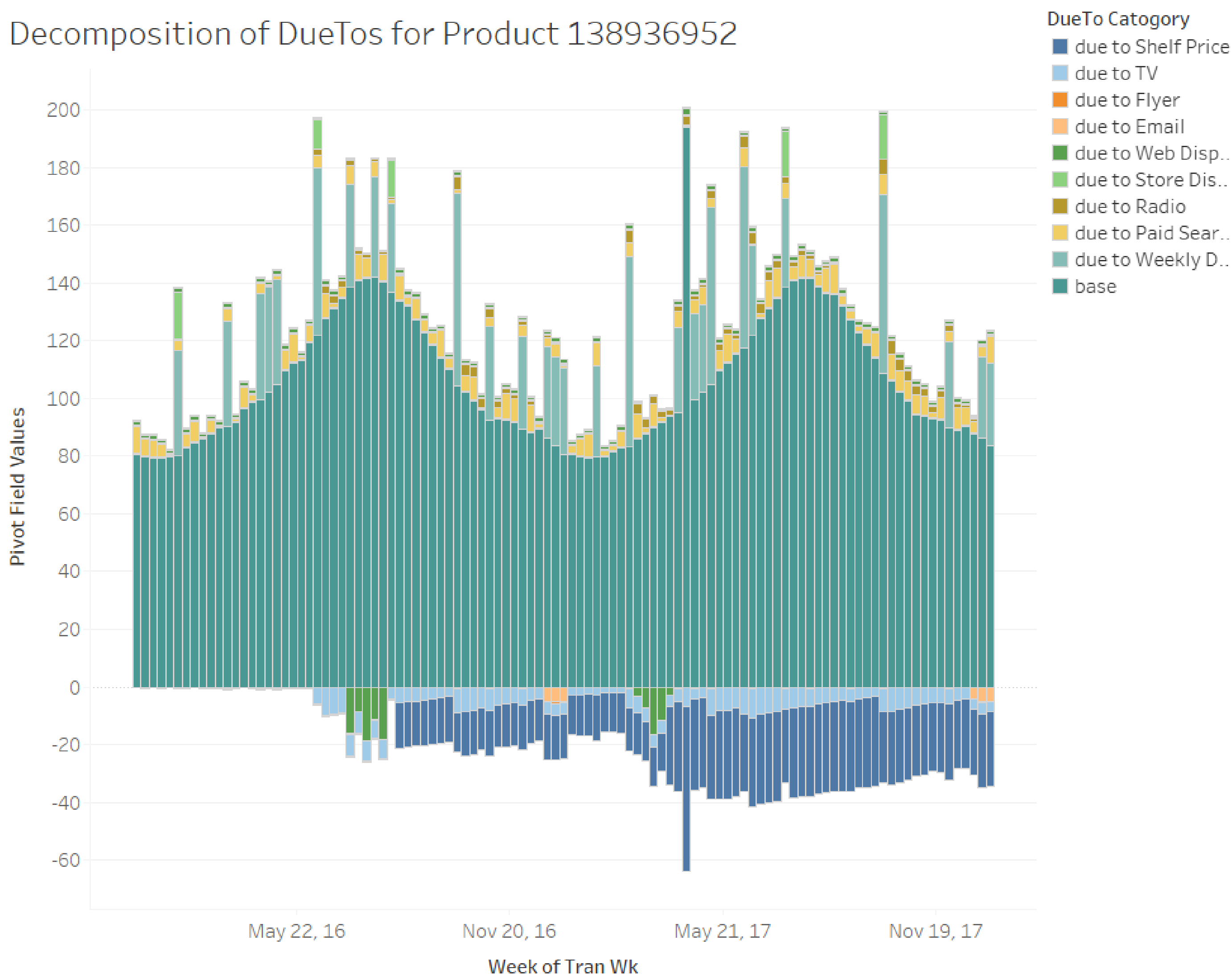
Next, we calculated the DueTos. The DueTos for each marketing vehicle is the difference between the predicted value of the target variable when this marketing vehicle is added and the predicted value of the target variable when this marketing vehicle is at its base.

The DueTos of each market vehicles for each product are shown as below.

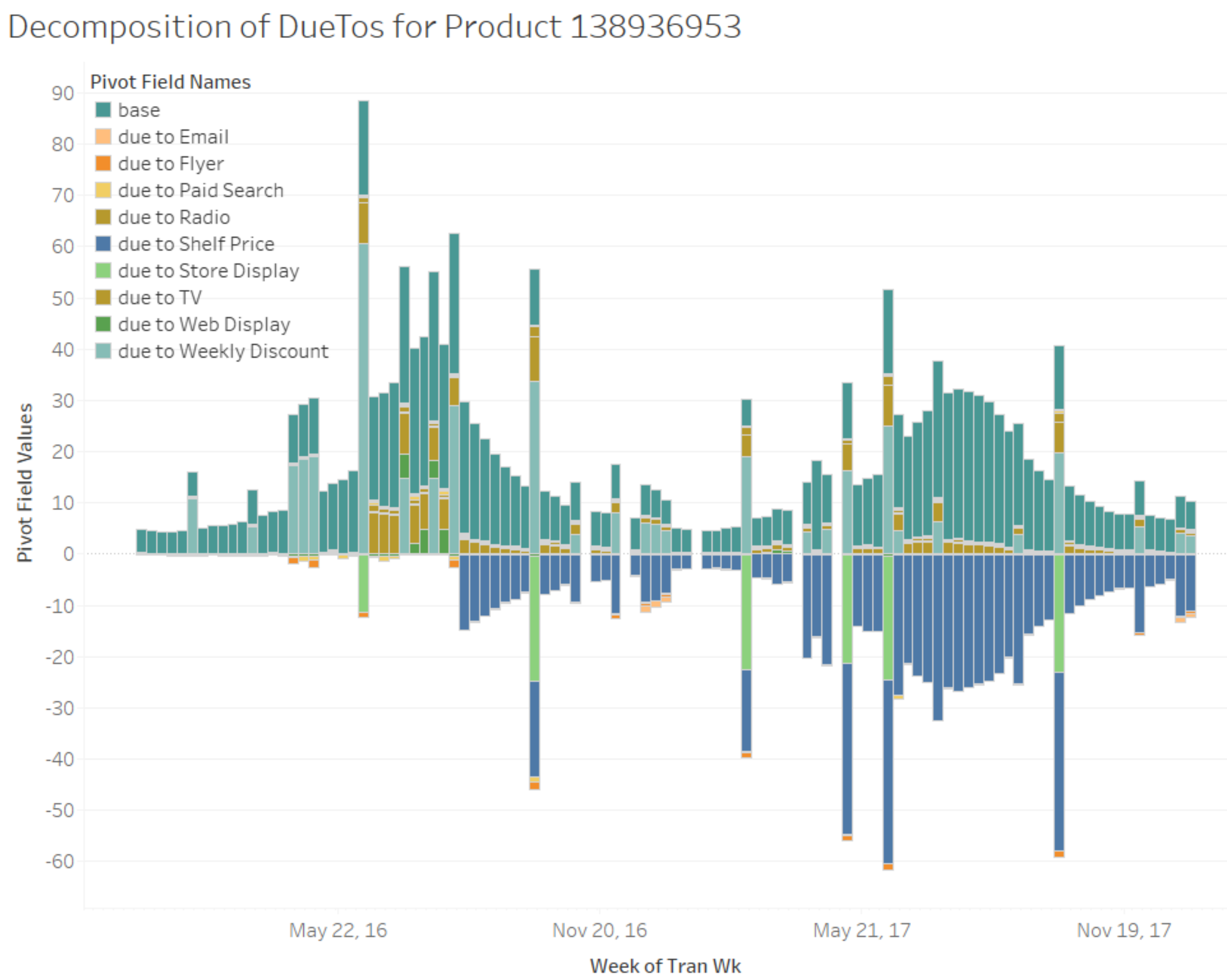
Product 138936951



Product 138936952



Product 138936953



Recommendation & Conclusions

Product: 138936951

Based on the decomposition of DueTos and summary of models for each product, we can tell that for product 138936951, ***TV*** is the most positive marketing approach since it shows the biggest lift in weekly sales quantity.

Radio has a negative impact; we would not recommend radio promotion activities for this product.

Product: 138936952

Based on the decomposition of DueTos and summary of models for each product, we can tell that for product 138936952, ***paid search*** has the most positive impact on weekly sales quantity. This product has a large base, discount has the most impact on the quantity. And the shelf price has a consistent effect on the quantity as well.

TV and web display have negative impact on the sales quantity.

Product: 138936953

Based on the decomposition of DueTos and summary of models for each product, we can tell that for product 138936953, *shelf price* has the most positive effect on sales quantity. This may because base forms a very small portion of the sales quantity. As for the marketing promotion vehicles, weekly discount has the most positive impact on quantity.

Shelf display has a negative impact on quantity.

Appendix

Model Summary for Product 138936951

```
> summary(lm1)

Call:
lm(formula = sales_tran ~ weekly_shelf_price + weekly_discount_rate +
    as.factor(Flyer) + Web_Display + Paid_Search + Email + TV +
    Radio + seas_index + as.factor(holiday), data = prod1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.99204 -0.24195  0.03903  0.23322  1.18027

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           3.02423    1.12974   2.677  0.00875 **
weekly_shelf_price   -4.05382    1.58225  -2.562  0.01198 *
weekly_discount_rate  2.57006    1.30702   1.966  0.05218 .
as.factor(Flyer)1     0.54754    0.36057   1.519  0.13220
Web_Display          -0.04823    0.05636  -0.856  0.39429
Paid_Search          -0.01163    0.05248  -0.222  0.82503
Email                 0.01775    0.05988   0.296  0.76758
TV                    2.17846    0.87823   2.481  0.01488 *
Radio                -1.37432    0.69870  -1.967  0.05211 .
seas_index            0.07900    0.07546   1.047  0.29783
as.factor(holiday)1  -0.15666    0.15986  -0.980  0.32960
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5055 on 95 degrees of freedom
Multiple R-squared:  0.3602,    Adjusted R-squared:  0.2929
F-statistic: 5.349 on 10 and 95 DF,  p-value: 3.114e-06
```

Model Summary for Product 138936952

```
> summary(lm2)

Call:
lm(formula = sales_tran ~ weekly_shelf_price + weekly_discount_rate + 
    as.factor(Flyer) + Web_Display + Paid_Search + Email + as.factor(Store_Display) + 
    TV + Radio + seas_index + as.factor(holiday), data = prod2)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.90369 -0.16656 -0.02567  0.18852  0.94810 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)               1305.55689  893.92026   1.460   0.1475    
weekly_shelf_price          -0.88668    0.18737  -4.732 7.82e-06 ***
weekly_discount_rate         4.96430    0.98809   5.024 2.40e-06 ***
as.factor(Flyer)1           -0.01011    0.12580  -0.080   0.9361    
Web_Display                 -0.07615    0.03893  -1.956   0.0534 .  
Paid_Search                742.92240  509.97910   1.457   0.1485    
Email                       -0.02185    0.03998  -0.547   0.5860    
as.factor(Store_Display)1    0.27829    0.18488   1.505   0.1356    
TV                          -0.33991    0.59977  -0.567   0.5722    
Radio                        0.15667    0.47437   0.330   0.7419    
seas_index                   0.34692    0.05243   6.616 2.24e-09 ***
as.factor(holiday)1         -0.03493    0.10609  -0.329   0.7427    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3459 on 94 degrees of freedom
Multiple R-squared:  0.6949,	Adjusted R-squared:  0.6591 
F-statistic: 19.46 on 11 and 94 DF,  p-value: < 2.2e-16
```

Model Summary for Product 138936953

```
> summary(lm3)

Call:
lm(formula = sales_tran ~ weekly_shelf_price + weekly_discount_rate + 
    as.factor(Flyer) + Web_Display + Paid_Search + Email + as.factor(Store_Display) + 
    TV + Radio + seas_index + as.factor(holiday), data = prod3)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.5305 -0.3967  0.0281  0.3350  1.0390 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)               10.73580    1.49431   7.184 1.98e-10 ***
weekly_shelf_price        -0.84920    0.09713  -8.743 1.28e-13 ***
weekly_discount_rate      10.13301    2.17744   4.654 1.13e-05 ***
as.factor(Flyer)1         -0.06837    0.29911  -0.229   0.8197    
Web_Display                0.05356    0.06721   0.797   0.4276    
Paid_Search                0.03746    0.06223   0.602   0.5488    
Email                     -0.03967    0.07291  -0.544   0.5877    
as.factor(Store_Display)1 -1.13884    0.55352  -2.057   0.0426 *  
TV                         0.90783    1.04685   0.867   0.3882    
Radio                      0.17042    0.83843   0.203   0.8394    
seas_index                 0.72719    0.09146   7.951 5.47e-12 ***
as.factor(holiday)1        0.02417    0.19325   0.125   0.9008    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5785 on 89 degrees of freedom
Multiple R-squared:  0.7977,	Adjusted R-squared:  0.7727 
F-statistic:  31.9 on 11 and 89 DF,  p-value: < 2.2e-16
```