Emory University

Master of Science in Business Analytics Program

Data-Driven Marketing Insights for Pernalonga

Alice Jiang, Cassio Salge, Qiyu Wang

MKT 680 Marketing Analytics

Professor David Sackin and Alvin Lim

2021/4/13

**Table of Contents**

Executive Summary…………………………………………………………………….3

1. Introduction………………………………………………………………………………3
   1. Background………………………………………………………………………3
   2. Business Context……………..………………………………………………...4
   3. Data Cleaning and Exploration………………………………………………..6
2. Modeling…………………………..…..………………………………………………....8
3. Model Evaluation.…………………………………………………………………........9
4. DueTos……………………...………………………………………………..………...13
5. Recommendations and Conclusion...……………………………………..………...17

**Executive Summary**

In this project, we aim to create models that break down the effect of various marketing vehicles for three Mahou San Miguel products based on their sale quantities at Pernalonga. After our analysis and modelling, we built a marketing mix model with recommendations for the three different products. We included the recommendations in the table below, as well as Mahou San Miguel’s current marketing strategies that we do not agree with. In our paper, we will provide context and details about how we came up with recommendations and conclusions. The table below summarizes our recommendations:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Product No.** | **Product** | **Continue** | **Might be considered** | **Discontinue** |
| 138936951 | Single Can | Discount  /TV(except winter) | Emails | Flyers/  /Radio  /WebDisplay  /PaidSearch |
| 138936952 | 6-pack | Discount/Radio | Paid Search | Flyers/Emails  /TV |
| 138936953 | Case | Discount  /TV(except winter)  /Radio(Holiday)  /WebDisplay(Summer) | Paid Search | Flyers/Emails |

**1.Introduction**

**1.1 Background**

Pernalonga is an undisputed leader in the retail space of Lunitunia, with over 10,000 products in 400+ categories. In order to drive sales, Pernalonga finds itself in a place where it depends on promotions, as over 30 percent of sales come from promotions. We've done a lot of research with data our client provided to figure out how their customers behave and react to promotions, so we're confident in our ability to push our analysis forward.

We know that Pernalonga can promote its products through a variety of marketing channels in order to increase key performance metrics like revenue and profit, and our task is to figure out how each of these channels contributes to the revenue change. As marketers, we must determine which channels, or marketing "vehicles," are most responsible for revenue and profit, and encourage Pernalonga to promote its products through the most efficient and effective channels.

In this report, we aim to analyze how different promotion vehicles contribute to Pernalonga’s success and to provide Pernalonga with data on how each type of promotion vehicle could affect its customers. We analyzed the effects of these vehicles on Mahou San Miguel’s (referred to as Mahou onwards) sales volume.

Mahou is a Spanish brewer whose three products are currently available in the Pernalonga stores. Mahou partners with Pernalonga extensively on promoting their products. Primary ways currently are to send out flyers or set up in-store displays. Mahou also employs other marketing vehicles such as email, web, TV and radio. To identify marketing vehicles and promotional opportunities, we built a marketing mix model that solves their question of increasing revenue through appropriate marketing vehicles.

**1.2 Business Context**

Mahou is in the leading position in the Spain beer market with about 36% of the market share. However, their products have not been as successful in Lunitunia and in the Pernalonga stores.

We found that there are three types of products from Mahou currently selling in the Pernalonga stores. These products are also sold differently in different packs. One is sold as a single can, another as a 6-pack and the third sold as a 24 pack. The 6-pack takes up 56% of Mahou’s total sales, which is the most out of the three packaging.

In order to make the marketing of these products successful, each type of packaging should target different consumers, and thus, different marketing vehicles. In the historical data, we noticed, the 24-pack sells at the highest prices, and accounts for a small portion of the margin. Many more single cans or 6-packs are sold compared to the sales quantity of the 24-pack which make them more profitable.

In the table below, we listed the average unit price, the total quantity sold, and most importantly, the discount percentage. The discount percentage is more relevant than the discount amount because it shows the relevancy of discount amount to unit price. As we see, the case package tends to have the highest average discount percentage while the single can has the least at 4%.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Product ID** | **Product Type** | **Unit Price (mean)** | **Price per Can (mean)** | **Total Quantity Sold** | **Total Sales** | **Discount Amount (mean)** | **Discount percentage**  **(mean)** |
| 138936951 | Single Can | $0.75 | $0.75 | 22,834 | $17,028.66 | $0.03 | 4% |
| 138936952 | 6-pack | $4.29 | $0.72 | 11,120 | $47,637.90 | $0.21 | 5% |
| 138936953 | Case | $15.64 | $0.65 | 1,276 | $19,957.74 | $1.22 | 8% |

We also analyzed seasonality because national holidays such as Christmas, even weekends could influence the volume sold for beers. We included the weekly seasonality data into our model to account for these fluctuations in efforts to predict the weekly sales volume.

Seasonality has a clear effect on the volume of beer sold at Pernalonga's stores, in both the beer category and Mahou products. June and July are when beer is sold the most, possibly because of the hot weather while November and December are sold least. Based on this seasonality, Pernalonga should consider spending a large portion of their marketing budget for the summer months when demand is highest.

Only six promotional channels have historically been used to promote all three of Mahou's products, according to the marketing vehicles data: TV, radio, paid search, web display, flyers, and emails. Mahou has only used store displays to promote its 6-pack and Case products, never the Single Can. Paid Search and Web Displays are measured in impressions, while TV and Radio are measured in Gross Rating Points, or GRP, which is defined as the total number of impressions of an advertisement with respect to a target audience. GRP is the most widely used metric for evaluating advertisements, and the data used to calculate it can be gathered in a variety of ways.

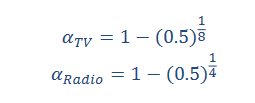
As mentioned before, we are building a predicting model based on weekly seasonality data. We will break down the sales quantity for Mahou’s products at Pernalonga into the amount that can be attributed to each of these vehicles using regression modeling techniques.

**1.3 Data Cleaning and Exploration**

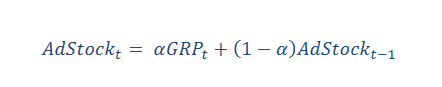
Firstly, we combined the two datasets to create a single file with all product and transaction data. This resulted in a dataset with nearly 30 million observations spread across 421 stores and 429 different product categories. There are 10,767 different products in our database. We know there are some discrepancies in the data that need to be addressed based on our previous analysis. The fact that there are only 753 unique transaction IDs assigned to the data is one of the most significant errors in the data. To account for the ID errors, we used a combination of customer ID, store ID, and transaction date to define the transaction IDs correctly, resulting in 2.83 million unique transactions with varying numbers of products purchased across all consumers. We also removed transactions from the final week of 2015 because there was only one day in this week with any transactions, skewing any results when aggregated at the weekly level.

Conversion of GRP to Reach for Radio & TV

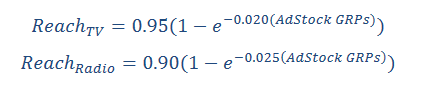
Before the model building process, we needed to do a few conversions in order to be successful in converting GRPs to Reach for each marketing vehicle. We first need to convert DROp to AdStock GRPs, which could then be converted to Reach. We calculated the value of alpha, decay parameter with the formula below and with information on TV advertisements have an 8-week half life and Radio advertisements have a 4-week half life.



The AdStock GRPs for the period impacted by the TV and Radio marketing vehicles were then calculated. The decay parameter and GRP are used to calculate AdStock GRPs, which help us account for the natural short-term retention of media effects on consumers over time.



Then, we converted AdStock GRPs to Reach. We know that Reach is a better indicator of the true impact of TV and radio ads, so we converted the AdStock GRPs to a 2+ Reach representation. This is the percentage of the target audience who has seen the advertisement at least twice in our case. Instead of using raw GRPs or impressions, we can model the percent of the target audience that is expected to retain the message over time when we use Reach instead of GRP. For TV and radio, the following equations were used to calculate reach.



It's also worth noting that some marketing channels were only used for a limited amount of time, so we considered the marketing vehicle's effect to be zero during those weeks. Because TV and radio advertisements were both first distributed in 2016, the sales volume in the weeks preceding June 5th, 2016 is not considered to be influenced by TV and radio advertisements. Furthermore, not every marketing vehicle is used for all three products . At any given time, flyers and store displays were only used for one or two Mahou products. We decided to build models separately for each of the three Mahou San Miguel products in order to uncover the effects of various marketing factors on different products.

Use of Other Marketing Vehicles’ data

TV and Radio advertisements needed the GRP to Reach conversion. However, Paid Search, Web Display, Email, Store Display, and Flyer, did not require any conversion. In the case of email, it was expressed as the number of times the email was circulated weekly. The number of impressions in the corresponding week was used for Paid Search and Web Display. These were kept as binary measures for Flyer and Store Display, with a 1 indicating that the promotional vehicle was used during the week and a 0 indicating that it was not.

One thing we noticed about both Email and Web Display is that the data values are extremely limited. There is only one unique value for circulations in email, and four unique values for impressions in web display. We debated whether these two variables should be treated as factors, but in the end, we decided to treat them both as integers because circulations and impressions could realistically be any whole number. This assumption is made in order to capture the nuances and intricacies of these variables as accurately as possible.

Because the weekly sales quantity of each product is our target variable, we feature engineered two product-related variables to use as independent variables for our model:

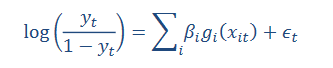
**Shelf Price** - for each product, we calculated the weighted average shelf price for each week. The weighted average shelf price is calculated by multiplying each store's average shelf price by a weight, in this case, total sales. Surprisingly, we discovered that the weekly weighted average price is the same as the average shelf price. However, the price of a product will not change across stores during a given week.

**Discount Amount** - for each product, we used the average discount rate for that week. The average discount for each product in each week is determined by multiplying the total discount amount by the total sales volume for that week.

Finally, to properly reflect the base of sales quantity that would not be affected by the addition of any marketing vehicles, we included seasonality for each week as well as a holiday index in our analysis. We used the data we were given as is for holidays, which indicated a week when a holiday occurred and which holiday it was. We wanted to see how a specific holiday affected sales volume, so we treated each holiday as a separate factor in the model to see if there were any differences between weeks with specific holidays and weeks without. In addition, we joined the seasonality index for Mahou’s products with the corresponding weeks.

**2. Modeling**

Our model aims to reflect how each marketing vehicle affects the sales quantity. Since we want to capture the complex interaction between various marketing vehicles as well as bounding the model with a theoretical maximum weekly sales quantity, we used the logit-based model with the formula below:



Before building the model, we assumed a value 20 percent higher than the product's historical maximum since it could be possible that there is a huge fluctuation in the sales quantity rise.

In our models, all marketing vehicles are included as columns. The only exception is that the model of the single can product did not include Store Display data. Additionally, we included seasonality index and holidays in all models because they are statistically significant to our models. We also included a weighted Shelf price (defined before) and average discount percentage by week. We discovered that three of the holiday factors were highly correlated with the Email marketing vehicle variable after running our initial models, and this makes sense because email promotions only took place during the weeks surrounding Christmas and New Year's, resulting in multicollinearity between the variables. To make sure we can include Email in the model, we double-checked which of these three holidays have an impact on sales volume. This was confirmed by comparing the volume of sales during the holiday weeks to the weeks immediately preceding and following them, and we discovered that sales tend to spike during the weeks of "PrXMAS" and "NEWYEAR." We also created a second model that excluded Email to see if these two variables were statistically significant in predicting weekly sales volume. As a result, we excluded the“XMAS”holiday from our final models, but included “PrXMAS,” “NEWYEAR,” and Email.

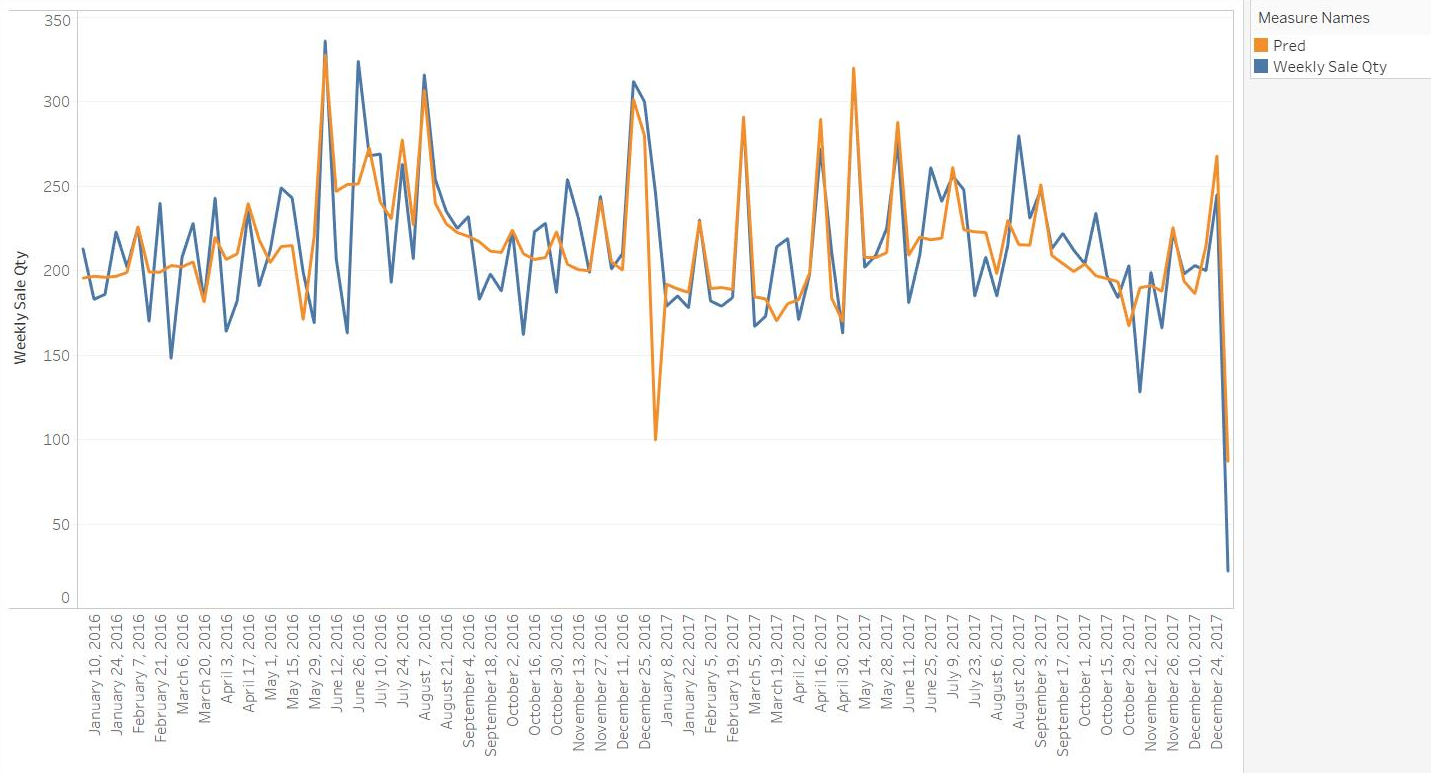
**3. Model Evaluation**

All three of our product-specific models were statistically significant with p-values less than 0.05, indicating that the models do not have a lack of fit issue, according to the F-test for Linear Regression. Variables like discount, some holidays, and the seasonality index seemed to have a positive impact on sales quantity for all products, whereas variables like some holidays seemed to have a negative impact. When higher discounts were offered, sales quantity increased. Shown in the table below, are our model evaluation metrics:

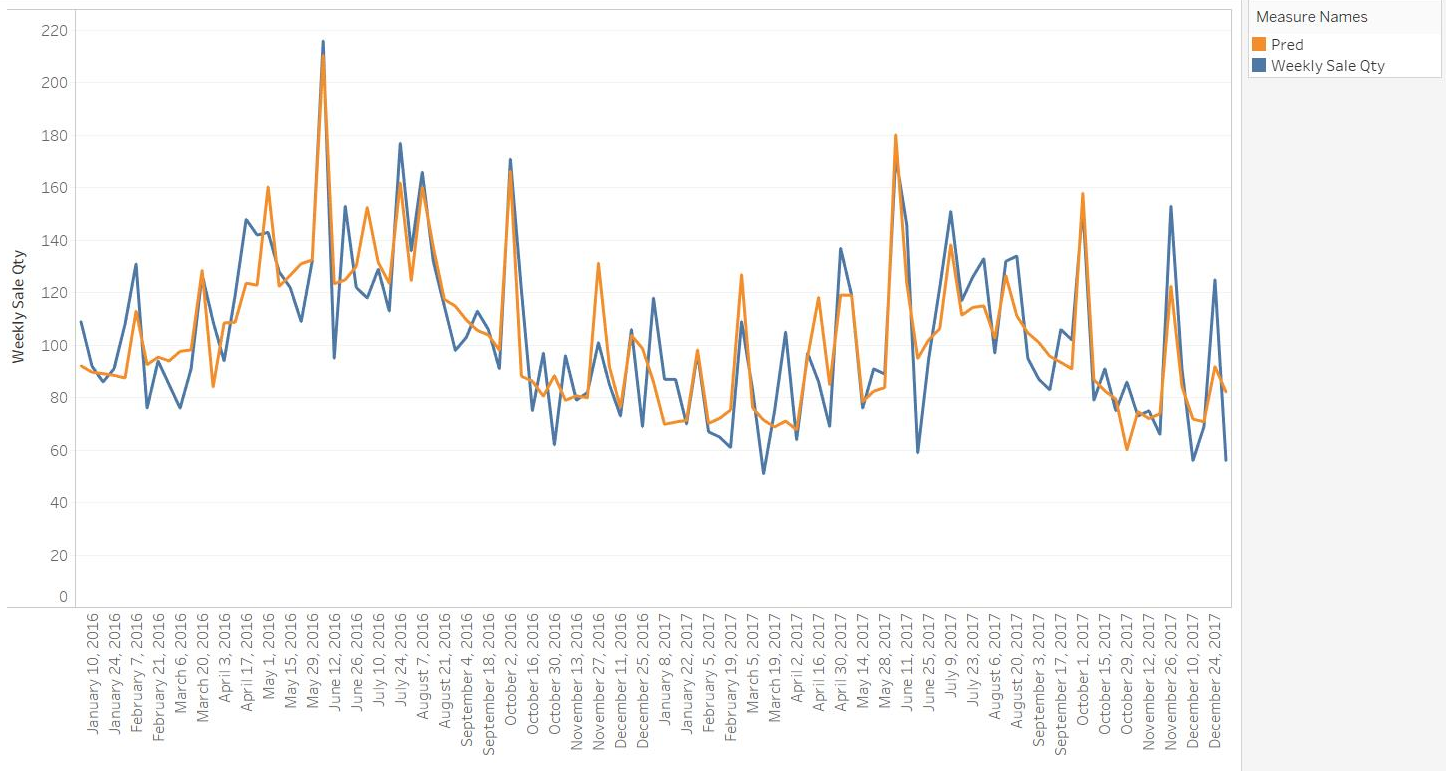
|  |  |  |  |
| --- | --- | --- | --- |
| **Product ID** | **RMSE** | **R2** | **Adjusted-R2** |
| 138936951  (Single Can) | 30.05 | 0.573 | 0.432 |
| 138936952  (6-pack) | 15.66 | 0.750 | 0.667 |
| 138936953  (Case) | 3.956 | 0.823 | 0.761 |

It's worth noting that, while Product 138936953 (Case) has the smallest RMSE and highest adjusted-R2 of the three product models. RMSE appears to be a good measure of model performance, along with adjusted-R2, which is a measure of the variance explained in quantities sold over time by the models, given the very different nature of the three products in terms of how their quantities are measured. Despite their differences, all three models did a good job of measuring the change in sales quantity over time. This is best illustrated by the line plots below, which show the actual weekly sales volume versus the predicted weekly sales volume for each model.

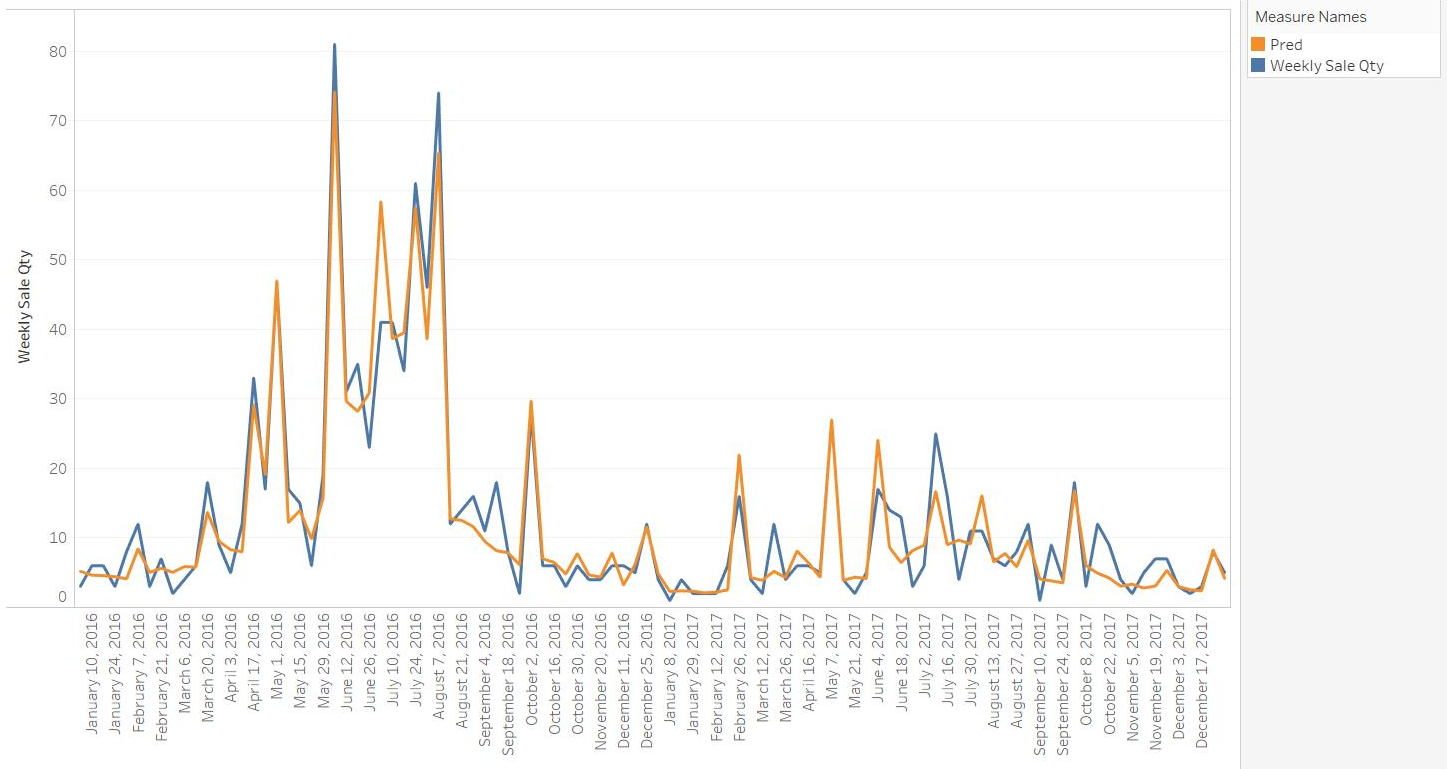
**Actual vs Predicted Sales for Product 138936951 (Single Can)**



**Actual vs Predicted Sales for Product 138936952 (6-pack)**



**Actual vs Predicted Sales for Product 138936953 (Case)**



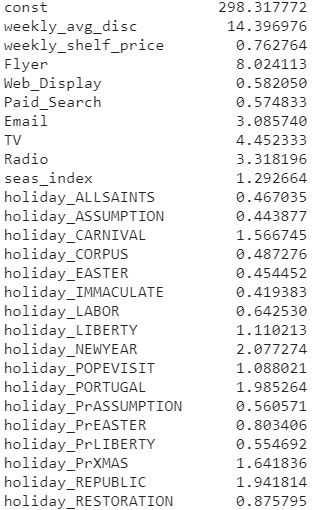
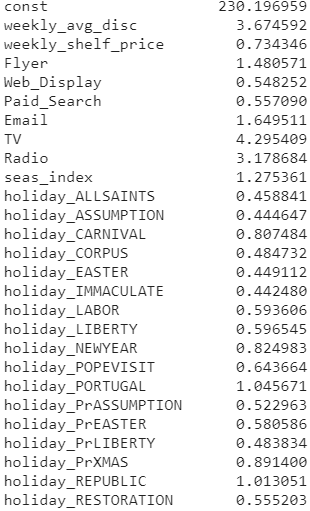
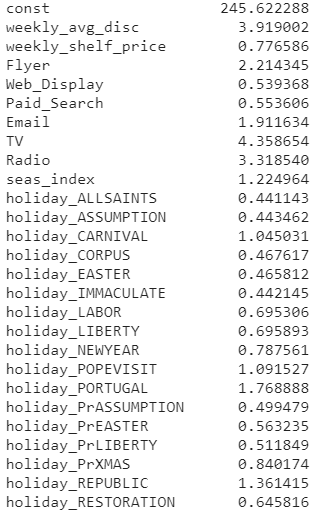
In general, all three models did a good job of detecting variations in sales volume. Given the data's time series nature, it was important to check for autocorrelation, which can cause a bias in the model coefficients' standard errors. We used the Durbin-Watson Statistic to determine this, and the results are shown below.

|  |  |
| --- | --- |
| **Product Type** | **Durbin-Watson Statistic** |
| Single Can | 1.89 |
| 6-pack | 2.18 |
| Case | 2.31 |

A value near 2 indicates a strong indication of no autocorrelation, which roughly corresponds to the values in our models. The p-value for all three models is greater than 0.05, indicating that the null hypothesis (H0= No first-order autocorrelation) cannot be rejected. As a result, the estimated standard errors should closely resemble the true standard errors.

Finally, we must ensure that the predictor variables in our models do not have any multicollinearity issues. To quantify the measure of multicollinearity, we used the Variance Inflation Factor, or VIF. In general, if VIF > 5, the variable is considered to have a high level of collinearity with other variables, indicating that the other variables can explain at least 80% of the variance of the variable. However, the VIF values for all of our products were less than 5, indicating that the model did not have any multicollinearity issues. Shown below is the VIF scores table for all three products:

138936951: 138936952: 138936953:



**4. DueTos**

The computation of DueTos is the final step in marketing mix modeling, and it is used to decompose sales quantity into the sales quantity Pernalonga would have seen regardless of any additional marketing activity, and the sales quantity due to each of Mahou San Miguel's various marketing initiatives. A marketing mix model's ultimate goal is to break down sales volume into its base and other components that can be attributed to each promotional vehicle. DueTos are the components that should be calculated to determine which promotional channel or vehicle is contributing the most to Mahou San Miguel's sales volume in Pernalonga's stores.

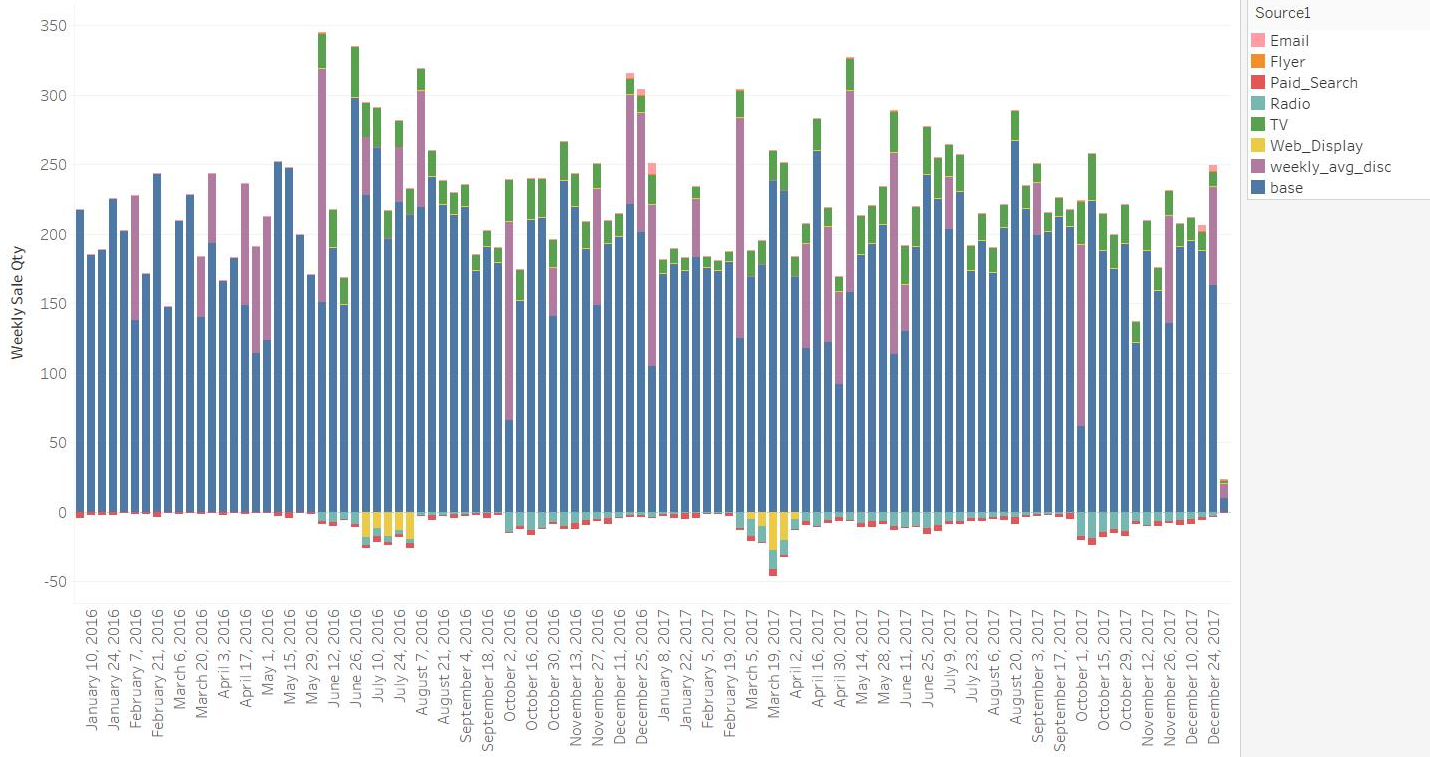
To begin, we needed to establish a sales quantity baseline. The portion of a marketing goal, in this case sales quantity, that is unaffected by the marketing vehicles of interest in the measurement is known as the base portion. In our case, the base for each product is the base value plus any seasonality and holiday contributions. The next step was to determine the starting point for each of our marketing vehicles. There are some heuristics we can use, such as setting the base equal to 0 in general for media and promotions. This includes any holidays or email promotions in our case. For price, the base can typically be set to either the product's historical average price or the first price it was sold at in our time frame, and in our case, we've chosen to set it to the first shelf price we see in the product's history.

We could start calculating each DueTo after we established the base values for the marketing vehicles for each of our three products. We can calculate the DueTo as the difference between the predicted value of the dependent variable when a marketing vehicle is added and the predicted value of the dependent variable when the marketing vehicle is kept at its base because we used a logit model, which is a type of multiplicative model. We calculated the predicted sales quantity of each product in each week using the actual variable values, and then we calculated the predicted sales quantity of the product using the base value for each marketing vehicle. Prior to any re-scaling, the difference between these values represents the corresponding DueTo for that marketing vehicle. This process was repeated for each of the three Mahou San Miguel products, with the data eventually being rescaled to match the original scale and units of our sales quantity variable, which is simply dollars.

Each of the predicted values in logit models has a transformation bias built in. We needed to scale each DueTo so that the sum of all DueTos equaled the corresponding product's actual weekly sales quantity in order to properly debias our model results. To do so, we added up all DueTos for each product in each week based on the predicted sales quantity for that week. We then divided the corresponding DueTo for each marketing vehicle by the sum of all DueTos in that week and multiplied it by the actual weekly sales quantity to get each re-scaled individual DueTo.

Each Mahou San Miguel product's weekly sales quantity was eventually broken down into the following components: Base (which includes price, seasonality and holidays), Discount, Paid Search, Web Display, Email, Store Display, Flyer, TV, and Radio. After scaling the results back to the original value of our sales quantity variable and appropriately debiasing the results, the following are the results and visual representations of each product's DueTo decomposition.

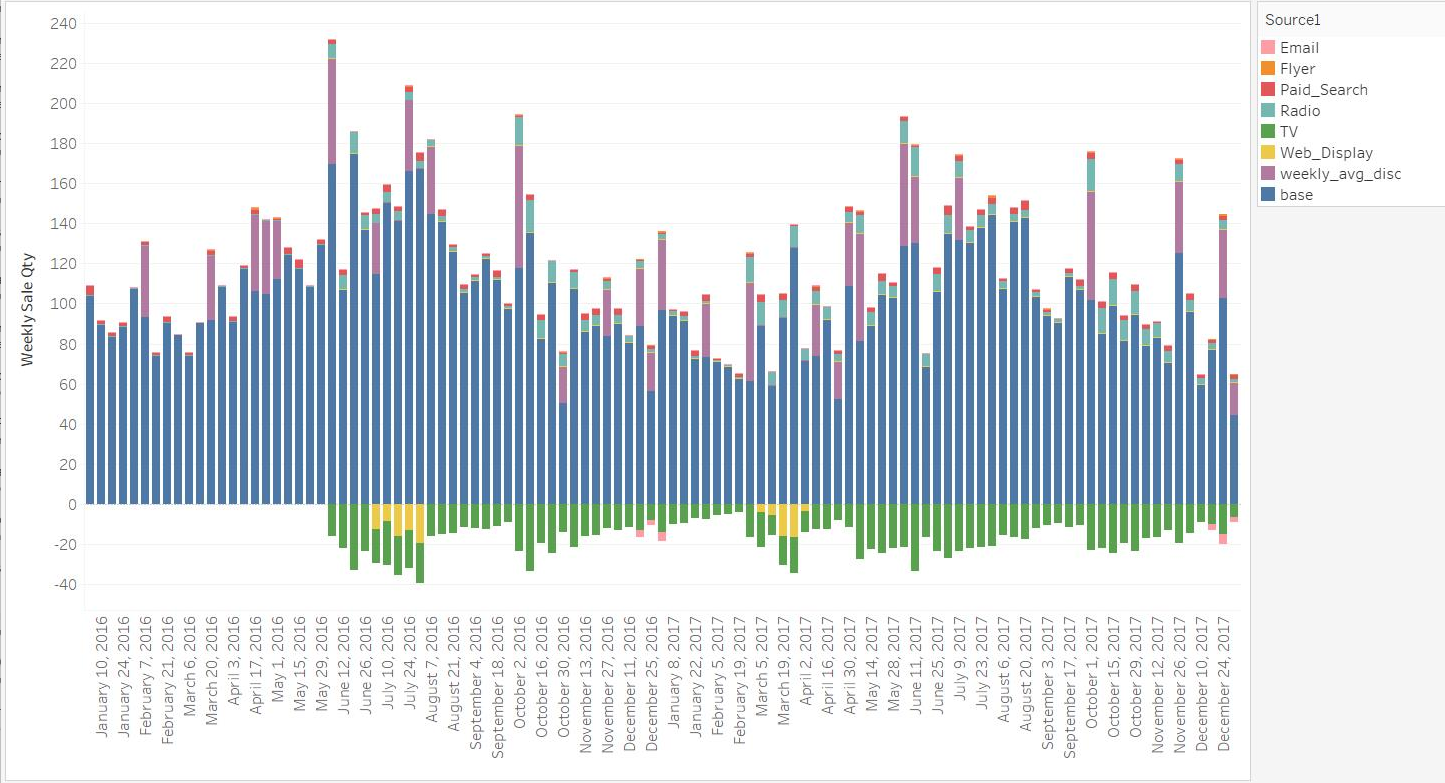
**DueTos for Product 138936951 (Single Can)**



The DueTo decomposition of the Single Can product that the additional positive contributions of marketing vehicles to sales quantity are primarily attributed to discounts and television, both of which appear all year. Discounts appear to contribute the most to sales volume in most weeks. During most weeks, Emails also make a positive contribution.

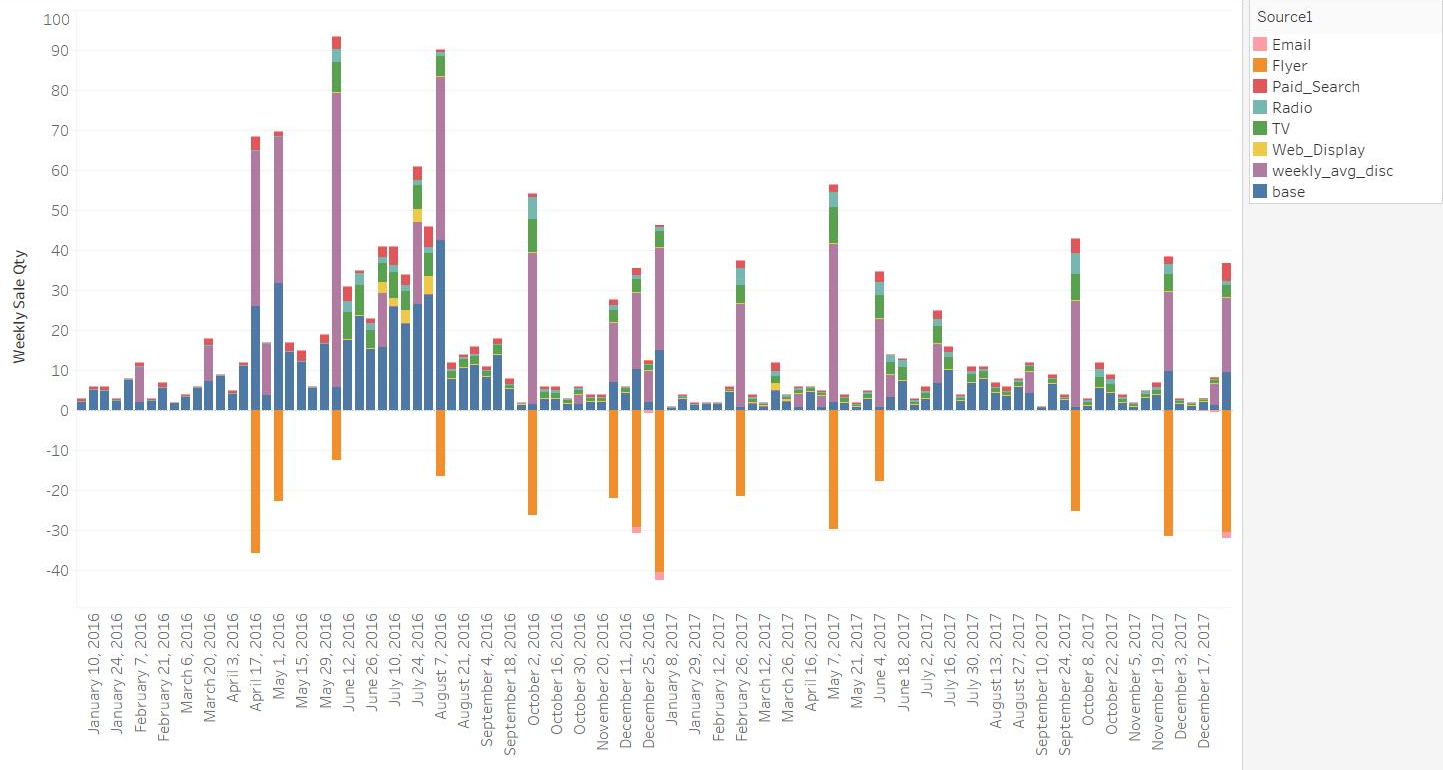
Also, we see what appears to be a negative effect from Web Displays in several of the weeks, as well as a negative effect from Radio in most weeks over the course of our data. This means that Radio is ineffective in selling Mahou San Miguel Single Cans in Pernalonga's stores. When we compare this to other traditional marketing methods, such as television, we can see that television appears to have a positive impact on Single Can sales volume. This is something to be cautious of, as a promotional event is unlikely to have had a negative impact on sales. It's more likely that sales volume fell at the same time as a promotional event, but concluding that the promotional event was the cause isn't always accurate. Mahou San Miguel and Pernalonga should look for more underlying reasons for the decrease in sales volume during these time periods in future models and further analysis.

**DueTos for Product 138936952 (6-pack)**



We see a different story for Mahou San Miguel 6-packs than we did for Single Cans. We see that Radio has a high positive effect on sales quantity for most weeks, while TV appears to have the biggest negative effect, with Web Displays and Emails having a negative effect on sales quantity for several weeks as well. We also see smaller but still significant positive effects from paid search and minor effects from Flyers. We see the most additions on top of the base during the winter months, but there are surprisingly fewer total sales during those months. The 6-packs appear to be more popular during the summer months, when people may be purchasing them to share with friends while the weather is nice.

**DueTos for Product 138936953 (Case)**



Finally, we looked at Mahou San Miguel's Case product's DueTo decomposition. We see much lower contributions to sales quantity across all marketing channels, with Flyers having a particularly negative impact. This product appears to have a seasonal sales pattern, with peaks during the summer months and troughs around the holidays. We see what appears to be large positive effects from discounts, just like the other two products. TV, Radio, Web Display, Paid Search appears to all have a positive impact on sales volume, where Web Display’s impact is seasonal.

**5. Recommendations and Conclusion**

We can conclude, based on everything we've learned from the sales quantity decomposition process, that there are some Mahou San Miguel promotional campaigns that are extremely effective at driving sales. However, we discovered that some promotional techniques have no discernible effect on driving incremental sales. As a result, according to our DueTos analysis, Mahou San Miguel should focus on the marketing vehicles that appear to drive the highest incremental sales for each product.

For this particular case, we don’t have information regarding the cost of each marketing vehicle, since the case was back in time around 2016 and not in the US, the price levels information online of marketing vehicles can be way off, especially when retail chains usually got better prices than the prices for the public. We would only make recommendations of continuation on the ones with significant positive impacts.

Discounts are the most effective vehicle for all three products and should be definitely continued, and an increase of investment in Discounts could also be considered. We see the effects of Discounts across all the places, which might indicate the purchasing behaviors are largely determined when customers are in-store.

TV has been very effective at driving incremental sales for both Cans and Cases for times except winter, and it may be possible to consider using this media for a longer period of time as well. However, it’s not effective for 6-packs and definitely should be removed from 6-packs marketing-mix planning for the future.

Radio media promotions should also be considered to increase reach as the second largest positive effects on 6-packs and cases. But Radio should be only considered for holidays for Cases. Paid Search is the other one that could be considered for 6-packs and cases for small amounts, since the return is positive but not significant.

We strongly advise that Flyers and Emails (except for Cans) be discontinued, as it does not appear to drive positive sales for most of time for all three products. Last but not the least, Web Display is only effective for Cases in Summer, and won’t provide positive effects for other times/products.

A/B testing across similar stores could be considered to validate the updated marketing mix planning, allowing Pernalonga and Mahou San Miguel to test the effectiveness of different marketing vehicles while controlling for other factors like price and discount.

Mahou San Miguel would be able to keep up with the marketing vehicles that drive the most incremental sales in the future according to our analysis.