Analysis and Pricing Recommendations for Beer Based on Past Sales Performance

Retail Analytics

Group 7

March 4, 2024

Executive Summary

The purpose of this retail analytics report is to examine how pricing and revenue for the products Old Style Lt Bonus 6(UPC7336097301), O'Doul's n/a Longneck (UPC1820000784), Budweiser Dry Beer(UPC1820000202), and Miller Genuine Draft(UPC3410017579) are related. We use a wide range of analytical techniques, such as decision trees, logistic regression, linear regression, and two-way analysis, to try and identify the different elements affecting the product's profits. The main goal is to pinpoint the most important element in the dynamic relationship between price and profits, opening the door for well-informed pricing tactics that maximize profits. Along with identifying the most profitable and lowest profit beers out of a variety of beers, we also examine the reasons behind its strong profitability to assist retailers in suggesting increased sales.

Introduction

In order to delve deeper into the retail analytics of BUDWEISER DRY BEER, a comprehensive examination of specific product UPCs is essential. This analysis will focus on three key UPCs associated with UPC1820000784, UPC7336097301, UPC1820000202, and UPC3410017579. By honing in on these unique product identifiers, our study aims to provide granular insights into the performance, distribution, and consumer behavior associated with these particular variants.

Figuring out the complex dynamics underlying these UPCs is essential to identifying market trends and patterns. Understanding the many variables that affect income, such as pricing changes, weekly variations, and store-specific effects, will be made easier with the use of machine learning and statistical models. By using this analytical method, we hope to provide decision-makers with useful information that they can use to improve overall business performance and optimize strategy in the highly competitive beverage market.

Date set Summary

Our analysis was based on data that was gathered by the University of Chicago Booth School of Business in collaboration with Dominick's Finer Foods, which is no longer in operation, from 1989 to 1994. The research focused on shelf management and pricing at the store level. The analysis of the price-profit relationship might be done with this data. Over 25 distinct categories are included in the 100-store chain of Dominick's Finer Foods in the Dominick dataset. We have decided to analyze beers out of all of them.

The beer dataset contains more than 3 million observations and 11 variables (store number, UPC, week, the number of actual items sold, the size of bundle, price, promotion, profit, the validity of data, profit_hex, and price_hex).

We drop the number of actual terms sold, equalling 0, which is move=0, and flagged data (ok=0) to make sure the data is valid and useable. Additionally, we recorded categorical variable sales. If the sale is a missing value, it recoded as 0; if the sale is “B,” indicating a bonus buy, it recoded as 1; if the sale is “S,” indicating a simple price reduction, it recoded as 2. Lastly, drop off missing values in the dataset.

```
     STORE            UPC                WEEK           MOVE             QTY        PRICE            SALE
 Min.   :  2.00   Min.   :2.940e+02   Min.   : 91.0   Min.   :  1.000   Min.   :1   Min.   : 0.010   Min.   :0.0000
 1st Qu.: 71.00   1st Qu.:3.410e+09   1st Qu.:160.0   1st Qu.:  2.000   1st Qu.:1   1st Qu.: 3.790   1st Qu.:0.0000
 Median :100.00   Median :6.207e+09   Median :227.0   Median :  3.000   Median :1   Median : 4.990   Median :0.0000
 Mean   : 90.49   Mean   :6.504e+09   Mean   :229.7   Mean   :  6.829   Mean   :1   Mean   : 5.689   Mean   :0.2818
 3rd Qu.:119.00   3rd Qu.:7.336e+09   3rd Qu.:295.0   3rd Qu.:  7.000   3rd Qu.:1   3rd Qu.: 7.050   3rd Qu.:1.0000
 Max.   :146.00   Max.   :7.971e+10   Max.   :399.0   Max.   :769.000   Max.   :2   Max.   :29.640   Max.   :2.0000
     PROFIT
 Min.   :-98.10
 1st Qu.:  7.83
 Median : 19.84
 Mean   : 16.35
 3rd Qu.: 24.09
 Max.   : 88.48
```

We are interested in 7 variables: store(store number), UPC(product number), week(time), move(quantity on sold), sale(promotion events), quantity(number of beers in the package), and profit(Gross Margin). The summary statistics provide an overview of the sales dataset, revealing the distribution of various variables related to beer sales. Stores vary from 2 to 146. There are more than 50 types of beer, and each type has a unique UPC. The data is from weeks 91 to 399, hinting at a dataset spanning multiple years. The maximum quantity sold is 769 and the minimum quantity sold is 1, and most beers are single-item purchases. The price fluctuates between $0.01 and $29.64. Finally, profit per transaction ranges from a loss of $98.10 to a gain of $88.48, indicating varied profit margins across sales.

Methodology

Part I

Beer Type determination

This part utilizes a combination of data manipulation and summarization methodologies facilitated by the ‘dplyr’ package in R to analyze profitability across different beer products identified by their Universal Product Code (UPC). Specifically, it involves grouping the dataset by UPC, calculating the total profit for each group, and then arranging these groups in descending or ascending order to identify the products with the highest and lowest profitability, respectively. This approach leverages descriptive analytics to summarize past sales data, offering insights into product performance that can inform strategic business decisions regarding inventory management, pricing strategies, and promotional efforts.

Part II

Simple Regression Model

The Simple Linear Regression (SLR) model is a fundamental statistical technique widely utilized for understanding the relationship between a single predictor variable and a response variable. The SLR model is applied to our dataset to analyze how individual predictors like price, week, move, and sale affect the continuous 'Profit' variable. This method allows us to understand the direct impact each variable has on profit, quantifying their relationship in a straightforward and efficient manner. By focusing on one predictor at a time, the model serves as an initial step in pinpointing significant variables for more in-depth analysis through complex modeling techniques.

Part III

Predictive modeling

Gradient Boosting Machines(GBM)

The Gradient Boosting Machines (GBM) model is a powerful machine learning technique used mainly for classification and regression approaches. The use of the GBM model in our dataset is to predict the 'Profit' variable based on several predictors, including price, week, move, and sale. Since the response variable in our dataset is continuous, the distribution parameter is set to 'Gaussian' to make it a regression prediction. In the model, 'n.trees' is initially set as 2000, 'interaction depth' as 3, and 'shrinkage' as 0.01. After selecting the most optimized model, we will try to change these initial values to determine any further changes in our measurements.

Neural Networks(NNET)

The Neural Network (NNET) model, applied for predicting the 'Profit' variable in our dataset, utilizes a complex architecture of interconnected layers and neurons to model the nonlinear relationships between predictors like price, week, move, and sale, and the profit outcome. This advanced methodology excels in identifying intricate patterns and interactions that simpler models might overlook, offering a nuanced understanding of factors influencing profitability. Key to the NNET model's application is the normalization of input variables and the careful selection of network parameters, with the model's predictive performance evaluated through metrics such as MAE, MSE, RMSE, and R-squared. By leveraging neural networks, we aim to derive actionable insights into optimizing pricing, inventory, and marketing strategies to enhance profit margins, showcasing the NNET model's potential to significantly contribute to strategic business decision-making.

Analysis

Part I - Part III

Analysis of the worst performance of Beer UPC 7336097301

In this section, we will focus on the beer type with the lowest overall profit. Our approach is to determine and predict what variable we can change to improve its profit or lessen its loss.

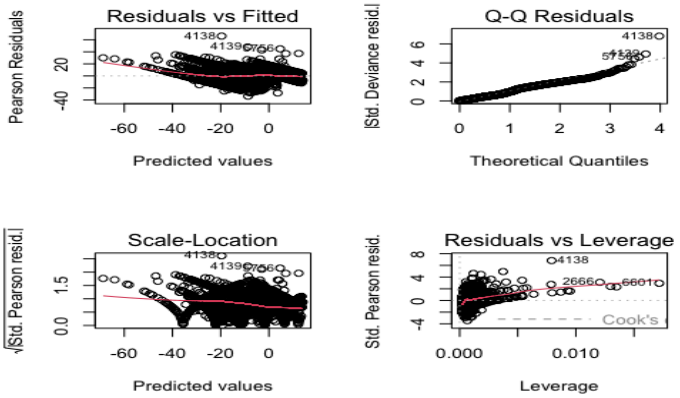
From Part I, we sum up all the profits in our dataset for every type of beer using the package "dplyr." The results show that the beer with the best profit through the year has a massive difference from different beers. The most profitable beer can earn 425692; the worst is losing 71977. Thus, our focus in this part of our report is to analyze the beer with UPC 7336097301, which brought a lot of loss. A short basic summary of data on the worst-performed beer shows that the price ranges from 7.49 to 11.99, and the gross margin ranges from -43.58 to 49.66. The whole summary can be found in the appendix at the end.

From Part II, we start analyzing the profit based on every variable shown in the dataset. Beginning with simple linear regression, we have applied linear regression to variables related to profit:

- Week (time)
- MOVE (quantity on sold)
- QTY (number of beers in the package)
- Price (retail price)

- Sale (events)
- Profit (gross margin).

Based on the result of the linear regression model for each of the variables above, we find that QTY doesn't have a tremendous relation to the change in profit. Hence, the QTY will not be considered as a significant variable for further determination. Other variables show their significance in predicting the value of profit; the R-squared containing this variable is 0.7354, which is not accurate enough to make the prediction. The following step is to apply logistic regression to the dataset. The result of the logistic regression model is similar to the simple linear regression, indicating applying a more complex model to predict the outcome.



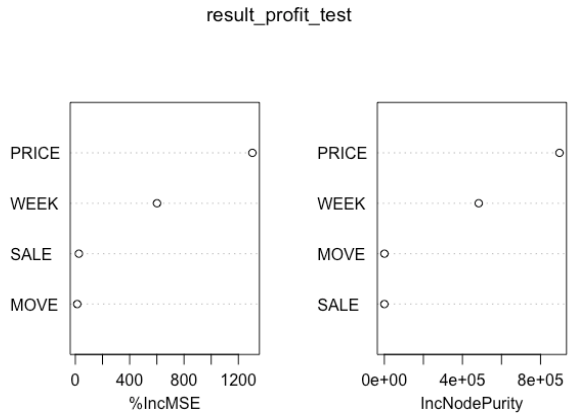
The appendix above shows how the logistic model fits our data.

Part III contains the use of several predicting models. In this part, we will use a decision tree(rpart), RandomForest, Gradient Boosting Machines(GBM), and multinomial logistic regression(NNET) to analyze our dataset. After applying our dataset to these models, we will predict using parts of our data and calculate the MAE, MSE, RMSE, and R-squared to find the most optimized model suitable for us to predict the optimal values for each variable. Here are the standards and their meanings that will help us choose from these models:

- MAE: offer a precise measurement of average error.
- MSE: give more weight to more significant errors when predicting errors
- RMSE: provide a balance by penalizing large errors while being in the same units as the response variable
- R-squared: indicates how much variance in the predicted model can be explained by the model built.

A lower MAE, MSE, and RMSE value will indicate better predictive accuracy. For R-squared, as it approaches 1, it would be considered superior in capturing the relationship between predictors and "Profit." After setting up the model and using the same test and train dataset as the predicted dataset, we can determine the errors in each model. In the RPART model, the MAE is 2.14, the MSE is 9.33, the RMSE is 3.05, and the R-squared is 0.9570. In the RandomForest model, the MAE is 0.05, the MSE is 0.18, the RMSE is 0.42, and the R-squared is 0.9991. In the GBM model, the MAE is 0.30, the MSE is 0.36, the RMSE is 0.60, and the R-squared is 0.9978. In the NNET model, the MAE is 2.91, the MSE is 19.67,

the RMSE is 4.44, and the R-squared is 0.8498. The result shows that the RandomForest model is most suitable for our model and provides an extremely high R-squared. As the figure shows, RandomForest indicates that price and week significantly influence the profit, while move and sale show less relative influence.



Lastly, to predict a wise way to improve the profit of this beer, we set up a test data consisting of each data for price, week, Move, and sale. By adjusting the values of these variables, we find two parts of data that show a positive profit for the beer. One selling situation is during week 177 and setting the retail price among 11; this situation will get a profit of 10 per Move. Another case is during weeks 260 to 300; setting the price at 20 can get a profit of 5.2 per Move. Both of the above will bring profits while selling. As the result shows, the beer has two selling times to earn profits. Other than this period, it is hard for the beer to earn profits; it suggests that this kind of beer is a seasonal product, which means stores should choose the season to sell it to earn profits on this product.

Part Ⅳ

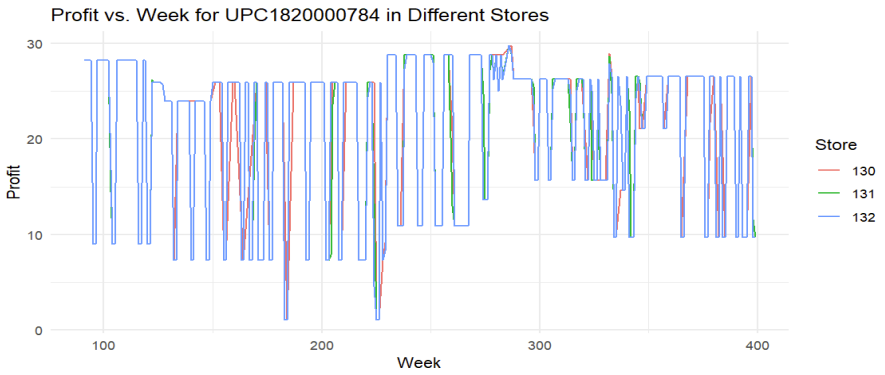
Analysis of the Historical Sales Performance of Beer UPC1820000784

In this section, we will select the best-performing beer of all, UPC1820000784, and analyze what characteristics it has exhibited over the past almost four hundred weeks of record sales. The criteria for selecting this beer was that UPC1820000784 helped the store achieve the greatest total revenue of all the beers.

After selecting the data for beer UPC1820000784, a linear regression was first fitted to its sales data. According to the summary results, the ***profit*** of UPC1820000784 is significantly correlated with the following three variables: ***price***, ***week***, ***store***, and ***sales***, with the p-value converging to 0. In addition, ***move*** is statistically significant from zero at the 1% level.

Next, the relationship between these variables will be visualized in the form of a plot.

(1)Sales performance of UPC1820000784 in different stores

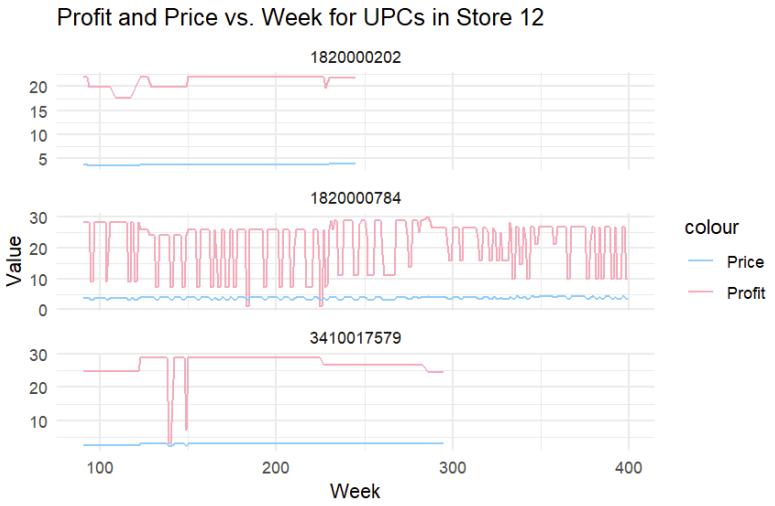


Profit vs. Week for UPC1820000784 in Stores 130, 131 and 132

Observation of the line graph shows that the trend in the number of UPC1820000784 sold in different stores is similar. In addition, the ***price*** and ***profit*** curves for UPC1820000784 at different stores roughly overlap (see Appendix for the figure). This indicates that the performance of UPC1820000784 is roughly stable across stores. This result conflicts with the results of the previous summary derived from the linear model - the p-value of the variable ***store*** tends to be zero, which should have indicated that the earnings of UPC1820000784 are very correlated with ***the store***. As can be seen from the figure, the reason for this divergence is that there is a significant difference in the amount of UPC1820000784 purchased by different stores, leading to a significant difference in the amount of UPC1820000784 sold, which ultimately and indirectly leads to a difference in profitability in the fitted results showing a correlation with the ***store***. However, in reality, the performance of UPC1820000784 is not significantly related to the ***store***.

This conclusion can be proved from another angle: when only the ***store*** is chosen as the independent variable to train the model, its p-value is much larger than 0.05 and tends to 1.

(2)Analysis of the sales performance of UPC1820000784 compared to other beers in store 12



Profit of UPC1820000784 Compared to Other Beers in Store 12 Over Time

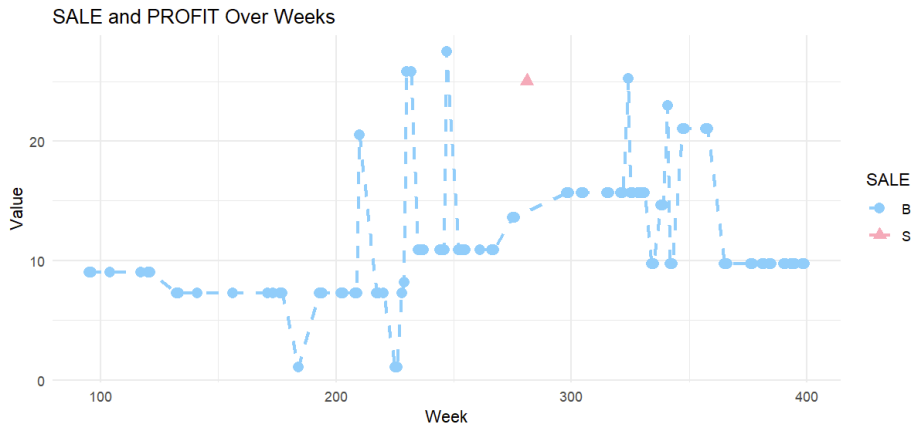
Observation of the line graph shows that UPC1820000784's earnings are more sensitive to changes in ***price*** than UPC1820000202. Whenever the price decreases, there is a significant decrease in the earnings of UPC1820000784. In terms of overall profitability, the highest profit of UPC1820000202 is around 20, while UPC1820000784 can reach 30, which is

significantly higher than that of UPC1820000202. Compared with UPC3410017579, the profit of UPC1820000784 is slightly less affected by price, and when the price decreases, the profit doesn't fall to almost zero like UPC1820000202. Instead, it still manages to stay around 10.

Compared to UPC1820000202, UPC1820000784 has a higher maximum profitability; compared to UPC3410017579, UPC1820000784 still maintains a certain level of profitability in case of higher prices and poses less risk for the store. Overall, UPC1820000784 performs well.

In addition, together with the ***Profit*** and ***Move*** over time curves and ***Price*** and ***Move*** over time curves for UPC1820000784 (see Appendix), the analysis shows that the reason why UPC1820000784's profitability decreases when the ***price*** is lowered is that when the ***price*** is lowered, the ***move(sales quantity)*** can be successfully stimulated though. However, because the pricing is too low, the ***profit*** that can be generated is very limited and even sometimes leads to negative profitability, i.e, the store, in order to sell the product, is selling it to the public at a price lower than the price of the product.

(3)Analysis of the Effects of Promotional Activities on the Sales of UPC1820000784



Profitability of UPC1820000784 by SALE

This previous analysis shows that the average profitability of UPC1820000784 is around 20. In this case, observing the effect of promotional activities on UPC1820000784 shows that the profitability of beer is generally lower than average when promotional activities are present. This suggests that promotional activities do not essentially help to improve the profitability of beer, and that promotional activities are more about clearing out stock and disposing of it so as to gain a small amount of benefit.

After the analysis above, it is clear that UPC1820000784, which has the largest total profit, has the following characteristics:

(1) Higher sensitivity of profit to changes in price.

(2) Similar sales performance in different stores.

(3) It can maintain a certain level of profitability even though it is lower due to price reduction.

(4) Profitability is generally lower than the usual level when there are promotional activities

In addition, in plotting the variation curve of the ***move*** with ***week*** for UPC1820000784, we expect to observe seasonal variations in beer sales. However, no clear trend was found.

After analyzing the beer above, it is clear that UPC1820000784 is a top-performing beer. It is profitable, popular with consumers, consistent across stores, and low risk. Therefore, from the point of view of helping stores maximize their benefits, we recommend that stores maintain a continuous stock of UPC1820000784. In addition, promotional activities do not seem to be a good thing as they reduce profitability. However, considering the reality that promotional activities are often related to clearance disposal and attracting new customers, they are also justified and necessary.

Conclusion

From the MAE, MSE, and R-square data, the prediction result of the RandomForest model has more accuracy compared to other models. According to the analysis, UPC7336097301 is a seasonal product that is most profitable during the designated selling times. It is difficult for the beer to turn a profit outside of these certain time ranges. Because of this seasonality, inventory management needs to be strategic in order to match product availability with the profitable selling seasons that have been recognized. It is recommended that stores sell UPC7336097301 beer at a price of 11 in the future, which will help maximize the store's benefits.

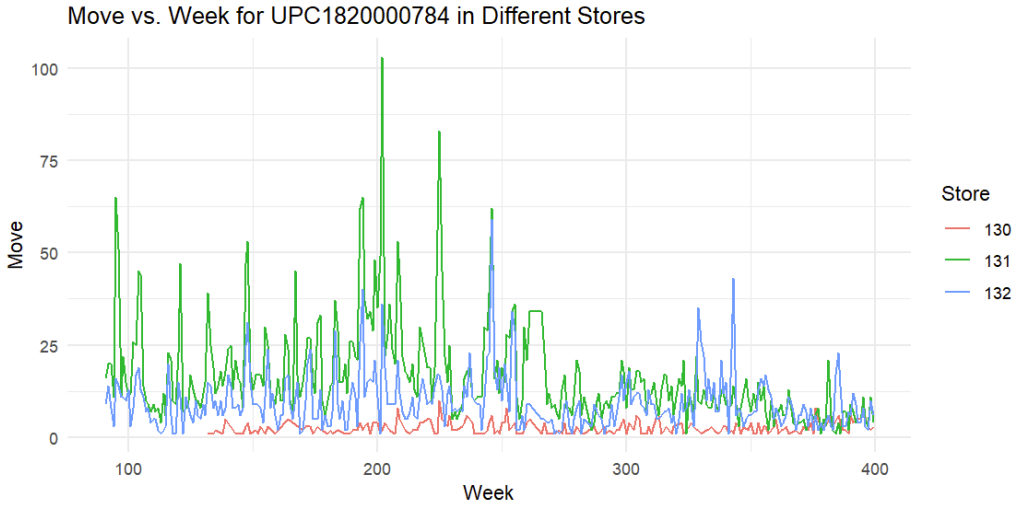
Based on the comprehensive study, UPC1820000784 is identified as a beer SKU that is popular with consumers, low-risk, and consistently profitable. It is advised that stores aim to optimize their gains by keeping a constant supply of UPC1820000784 on hand, considering its favorable performance metrics. The results also imply that since promotional efforts may have a transient negative effect on profitability, care should be taken when putting them into practice.

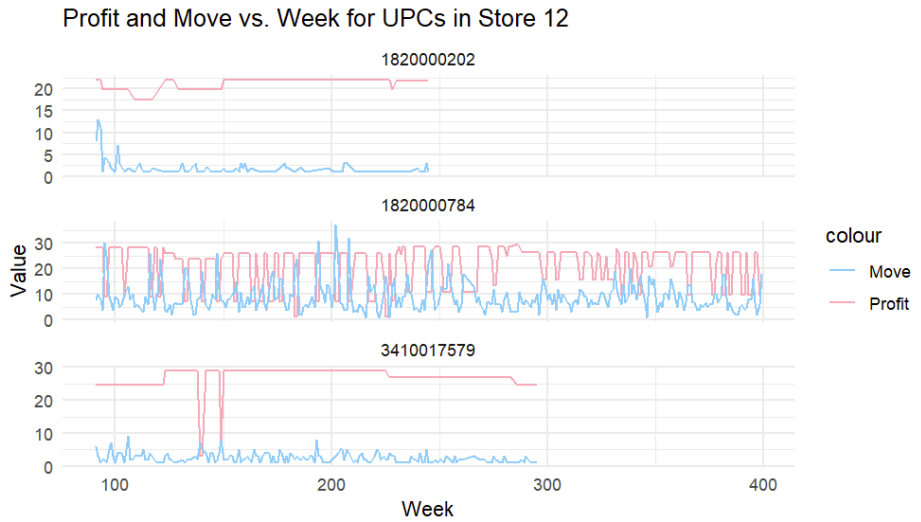
It is imperative to recognize that promotional endeavors, although they may have an impact on immediate financial gains, can be justifiably justified for objectives like client acquisition and clearance disposal. Consequently, it is advised to take a balanced strategy, balancing the advantages of promotional initiatives against UPC1820000784's steady profitability. To improve overall performance and customer satisfaction, stores can use this strategic perspective to help them optimize their inventory management and promotional methods.

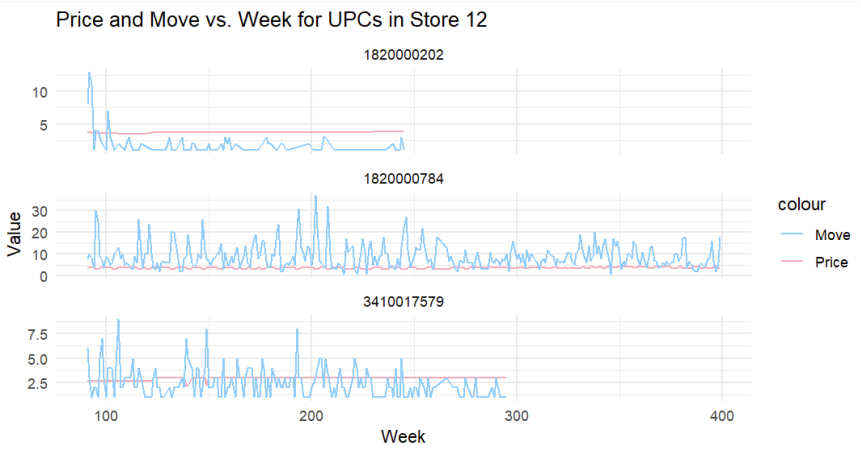
Reference

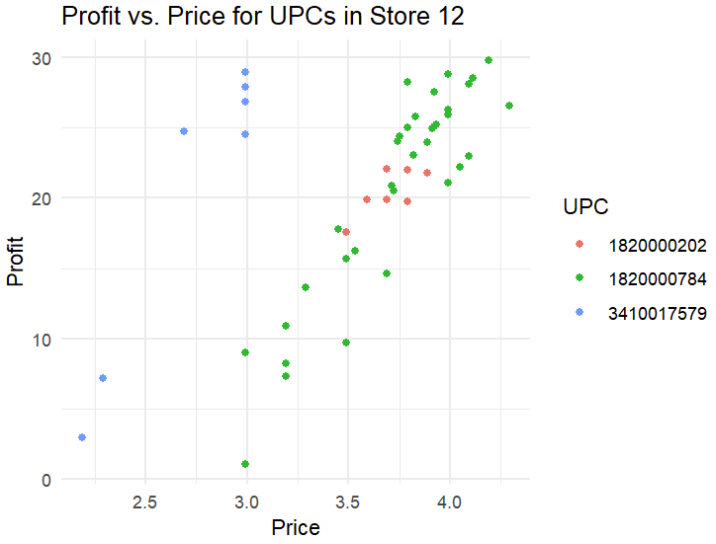
Chicago Booth School of Business. (n.d.). Dominick's dataset. Kilts Center for Marketing. Retrieved March 3, 2024, from https://www.chicagobooth.edu/research/kilts/research-data/dominicks

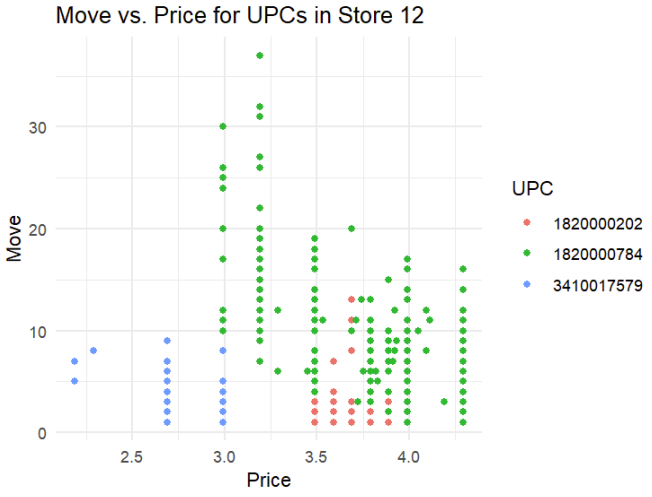
Additional Appendix

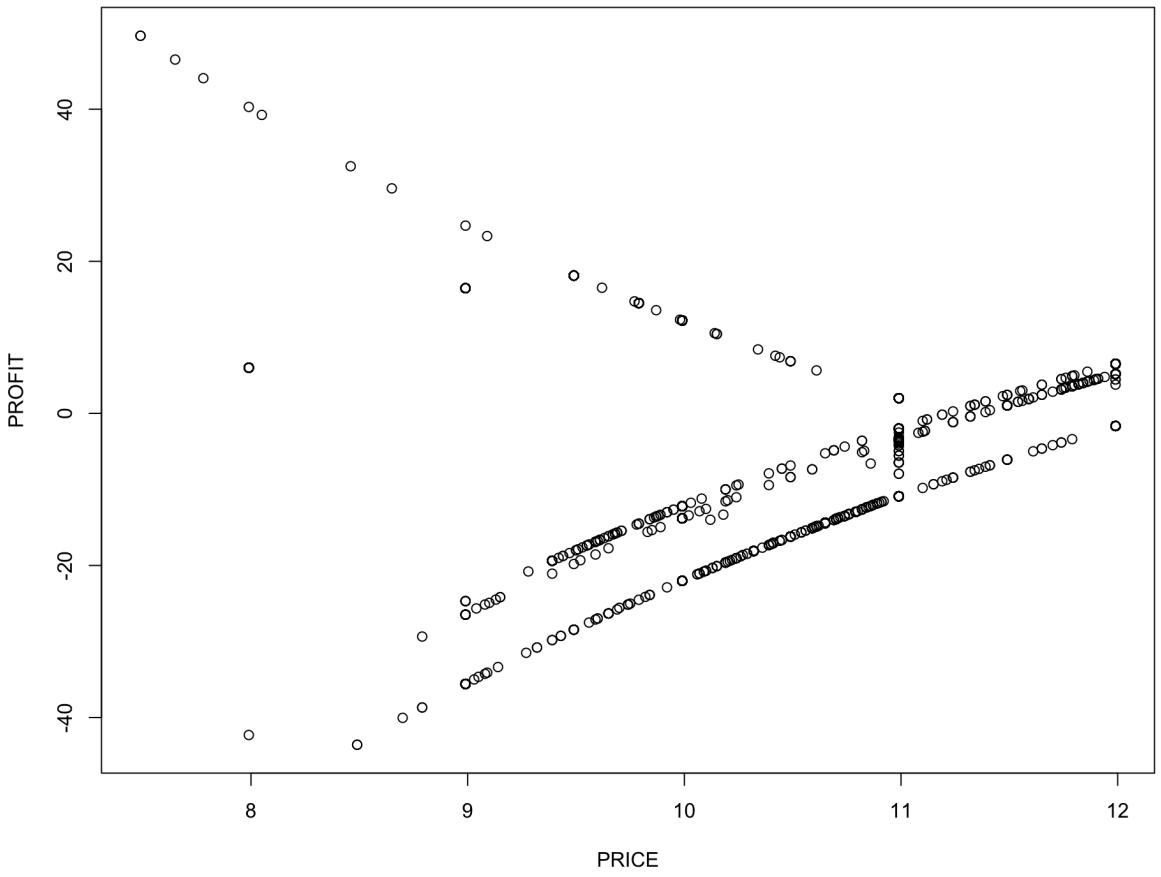





Price and Move vs. Week for UPCs in Store 12
1820000202
1820000784
3410017579
Value
Week
colour
Move
Price


Profit vs. Price for UPCs in Store 12
Profit
Price
UPC
1820000202
1820000784
3410017579


Move vs. Price for UPCs in Store 12
Move
Price
UPC
1820000202
1820000784
3410017579



Graph of Profit vs Price for all products

STORE	UPC	WEEK	MOVE	QTY	PRICE
Min. : 5.00	Min. :7.336e+09	Min. : 91.0	Min. : 1.000	Min. :1	Min. : 7.49
1st Qu.: 70.00	1st Qu.:7.336e+09	1st Qu.:193.0	1st Qu.: 2.000	1st Qu.:1	1st Qu.: 9.99
Median : 95.00	Median :7.336e+09	Median :242.0	Median : 5.000	Median :1	Median :10.99
Mean : 87.79	Mean :7.336e+09	Mean :230.6	Mean : 8.351	Mean :1	Mean :10.60
3rd Qu.:117.00	3rd Qu.:7.336e+09	3rd Qu.:273.0	3rd Qu.: 11.000	3rd Qu.:1	3rd Qu.:10.99
Max. :139.00	Max. :7.336e+09	Max. :317.0	Max. :106.000	Max. :1	Max. :11.99

SALE	PROFIT	OK	PRICE_HEX	PROFIT_HEX
Min. :0.0000	Min. :-43.580	Min. :1	Length:7277	Length:7277
1st Qu.:0.0000	1st Qu.:-22.020	1st Qu.:1	Class :character	Class :character
Median :1.0000	Median :-10.910	Median :1	Mode :character	Mode :character
Mean :0.5651	Mean : -9.891	Mean :1		
3rd Qu.:1.0000	3rd Qu.: 2.000	3rd Qu.:1		
Max. :2.0000	Max. : 49.660	Max. :1		

Summary data of the worst-performed beer