Orange Juice Sales at Wasatch Grocery Chain

Determining Factors and Predictive Modelling of Minute Maid Sales

Chris Gearheart and Chris Porter

2022-11-02

```r
set.seed(1234)
df <- read.csv(url("http://data.mishra.us/files/project/OJ_data.csv"))
df[-1] <- lapply(df[-1],as.numeric)
df$Purchase <- as.factor(df$Purchase)
datasplit <- createDataPartition(df$Purchase, p= 0.7, list=FALSE) # Documentation says this should equa
train <- df[datasplit, ]
test <- df[-datasplit, ]
```

Introduction

Wasatch Grocery Chain (WGC) is a regional grocery chain operating in the Intermoutain West of the US. WGC sells two brands of orange juice in its stores, Citrus Hill (CH) and Minute Maid (MM) of which MM is the more profitable to the company. This report will identify what customer factors within available data contribute to purchase of MM over CH, as well as to what degree these factors influence customer choice. In addition, a predictive model is created that will allow the Sales Department to identify other customers within our customer base that are more likely to purchase Minute Maid brand orange juice, thus driving profitability across the company.

Available Data

The data set used in this report contains 13 possible predictor variables as well as 1 outcome variable, Purchase, which records whether or not a customer purchased MM. There are a total of 1070 observations in the data set. The data set was further partitioned into a **training** data set, containing 750 observations, and a validation **testing** data set containing 320 observations.

```r
summary(df)
```

```
##  Purchase      PriceCH         PriceMM          DiscCH           DiscMM
##  0:417    Min.   :1.690   Min.   :1.690   Min.   :0.00000   Min.   :0.0000
##  1:653    1st Qu.:1.790   1st Qu.:1.990   1st Qu.:0.00000   1st Qu.:0.0000
##           Median :1.860   Median :2.090   Median :0.00000   Median :0.0000
##           Mean   :1.867   Mean   :2.085   Mean   :0.05186   Mean   :0.1234
##           3rd Qu.:1.990   3rd Qu.:2.180   3rd Qu.:0.00000   3rd Qu.:0.2300
##           Max.   :2.090   Max.   :2.290   Max.   :0.50000   Max.   :0.8000
##    SpecialCH        SpecialMM         LoyalCH          SalePriceMM
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000011   Min.   :1.190
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.325257   1st Qu.:1.690
##  Median :0.0000   Median :0.0000   Median :0.600000   Median :2.090
##  Mean   :0.1477   Mean   :0.1617   Mean   :0.565782   Mean   :1.962
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.850873   3rd Qu.:2.130
##  Max.   :1.0000   Max.   :1.0000   Max.   :0.999947   Max.   :2.290
##    SalePriceCH      PriceDiff        PctDiscMM        PctDiscCH
##  Min.   :1.390   Min.   :-0.6700   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:1.750   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :1.860   Median : 0.2300   Median :0.0000   Median :0.00000
##  Mean   :1.816   Mean   : 0.1465   Mean   :0.0593   Mean   :0.02731
##  3rd Qu.:1.890   3rd Qu.: 0.3200   3rd Qu.:0.1127   3rd Qu.:0.00000
##  Max.   :2.090   Max.   : 0.6400   Max.   :0.4020   Max.   :0.25269
##  ListPriceDiff
##  Min.   :0.000
##  1st Qu.:0.140
##  Median :0.240
##  Mean   :0.218
##  3rd Qu.:0.300
```

```
## Max.   :0.440
```

```
summary(test)
```

```
## Purchase    PriceCH         PriceMM          DiscCH             DiscMM
## 0:125   Min.   :1.69   Min.   :1.690   Min.   :0.00000   Min.   :0.0000
## 1:195   1st Qu.:1.79   1st Qu.:1.990   1st Qu.:0.00000   1st Qu.:0.0000
##         Median :1.86   Median :2.090   Median :0.00000   Median :0.0000
##         Mean   :1.87   Mean   :2.081   Mean   :0.05362   Mean   :0.1164
##         3rd Qu.:1.99   3rd Qu.:2.180   3rd Qu.:0.00000   3rd Qu.:0.2000
##         Max.   :2.09   Max.   :2.290   Max.   :0.50000   Max.   :0.8000
##   SpecialCH        SpecialMM         LoyalCH          SalePriceMM
## Min.   :0.0000   Min.   :0.0000   Min.   :0.000011   Min.   :1.190
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.374063   1st Qu.:1.690
## Median :0.0000   Median :0.0000   Median :0.600000   Median :2.090
## Mean   :0.1281   Mean   :0.1562   Mean   :0.578387   Mean   :1.965
## 3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.863765   3rd Qu.:2.130
## Max.   :1.0000   Max.   :1.0000   Max.   :0.999870   Max.   :2.290
##   SalePriceCH       PriceDiff        PctDiscMM         PctDiscCH
## Min.   :1.390   Min.   :-0.6700   Min.   :0.00000   Min.   :0.00000
## 1st Qu.:1.750   1st Qu.: 0.0000   1st Qu.:0.00000   1st Qu.:0.00000
## Median :1.860   Median : 0.2300   Median :0.00000   Median :0.00000
## Mean   :1.816   Mean   : 0.1485   Mean   :0.05578   Mean   :0.02831
## 3rd Qu.:1.890   3rd Qu.: 0.3000   3rd Qu.:0.10050   3rd Qu.:0.00000
## Max.   :2.090   Max.   : 0.6400   Max.   :0.40201   Max.   :0.25269
## ListPriceDiff
## Min.   :0.0000
## 1st Qu.:0.1400
## Median :0.2400
## Mean   :0.2113
## 3rd Qu.:0.2700
## Max.   :0.4400
```

```
summary(train) #need to equalize the 0/1 split in train data set
```

```
## Purchase    PriceCH          PriceMM          DiscCH             DiscMM
## 0:292   Min.   :1.690   Min.   :1.690   Min.   :0.00000   Min.   :0.0000
## 1:458   1st Qu.:1.790   1st Qu.:2.090   1st Qu.:0.00000   1st Qu.:0.0000
##         Median :1.860   Median :2.110   Median :0.00000   Median :0.0000
##         Mean   :1.866   Mean   :2.087   Mean   :0.05111   Mean   :0.1263
##         3rd Qu.:1.990   3rd Qu.:2.180   3rd Qu.:0.00000   3rd Qu.:0.2400
##         Max.   :2.090   Max.   :2.290   Max.   :0.50000   Max.   :0.8000
##   SpecialCH       SpecialMM        LoyalCH          SalePriceMM
## Min.   :0.000   Min.   :0.000   Min.   :0.000014   Min.   :1.190
## 1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.320000   1st Qu.:1.690
## Median :0.000   Median :0.000   Median :0.587000   Median :2.090
## Mean   :0.156   Mean   :0.164   Mean   :0.560404   Mean   :1.961
## 3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.841445   3rd Qu.:2.130
## Max.   :1.000   Max.   :1.000   Max.   :0.999947   Max.   :2.290
##   SalePriceCH       PriceDiff        PctDiscMM         PctDiscCH
## Min.   :1.390   Min.   :-0.6700   Min.   :0.0000   Min.   :0.00000
## 1st Qu.:1.750   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.:0.00000
## Median :1.860   Median : 0.2400   Median :0.0000   Median :0.00000
## Mean   :1.815   Mean   : 0.1456   Mean   :0.0608   Mean   :0.02689
## 3rd Qu.:1.890   3rd Qu.: 0.3200   3rd Qu.:0.1183   3rd Qu.:0.00000
```

```
##  Max.   :2.090   Max.   : 0.6400   Max.   :0.4020   Max.    :0.25269
##  ListPriceDiff
##  Min.   :0.0000
##  1st Qu.:0.1400
##  Median :0.2400
##  Mean   :0.2209
##  3rd Qu.:0.3000
##  Max.   :0.4400
```

Methods

Results

Conclusions and Recommendations

Brand

Sales