TECA Exploratory Analysis

Aishat

2023-06-21

TECA owns over 150 convenience stores and gas stations throughout the middle of the United States. TECA has been attentive to the trends and changes in their marketplace, the political landscape and the world around them. And they see that demand for fossil fuels, such as the natural gas that's eventually used to run cars is going to continue to decrease. This is a problem for TECA and owner of gas stations and convenience stores because selling gas for cars and trucks is a huge part of their business. If more and more cars are electric and fewer and fewer cars run on gas, a major portion of TECA's revenues will dry up.

Due to this reason, TECA is focused on increasing the sale of their high gross profit margin products, these products have high revenue relative to what it costs to sell them. Some examples of TECA's highest gross profit margin products are pizza cold sandwiches, cold dispense beverages, breakfast sandwiches, and hot dispense beverages. They would like to be able to predict when and why a transaction is to be for a product that's high in gross profit margin. If they can get some answers to this question, they can start doing even more to prepare for the uncertain future.

```
# Installing necessary packages
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
install.packages("caret")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
install.packages("class")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
```

```
## (as 'lib' is unspecified)
```

```
#loading the installed packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(class)
```

```
# Reading the dataset
knn_input <- readRDS("HL4_ELu6R2a-PxC7uqdmpA_bfd52f5a7dcd472ca67b2ac393441ef1_knn_input.rds")
```

```
# Checking the structure of the dataset
str(knn_input)
```

```
## 'data.frame':    20000 obs. of  38 variables:
##  $ high_gpm                  : Factor w/ 2 levels "low","high": 2 1 1 2 2 2 1 1 1 1 ...
##  $ revenue                   : num  0.99 27.25 2.49 2.26 3 ...
##  $ quarter.1                 : int  1 0 1 1 0 0 0 0 0 0 ...
##  $ quarter.2                 : int  0 1 0 0 1 1 1 0 0 0 ...
##  $ quarter.3                 : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ quarter.4                 : int  0 0 0 0 0 0 0 1 0 1 ...
##  $ income                    : int  65268 58854 61115 45874 43547 41827 45874 66250 70963 61115 ...
##  $ bachelors_degree          : int  118 68 264 2488 502 169 2488 22 1959 264 ...
##  $ population                : int  1003 734 2048 30327 10615 2379 30327 567 16052 2048 ...
##  $ state_province.Alabama    : int  0 0 0 0 1 0 0 0 1 0 ...
##  $ state_province.Arkansas   : int  0 0 0 1 0 0 1 0 0 0 ...
##  $ state_province.Colorado   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ state_province.Iowa       : int  1 0 0 0 0 0 0 1 0 0 ...
##  $ state_province.Minnesota  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ state_province.Missouri   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ state_province.Nebraska   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ state_province.Oklahoma   : int  0 0 1 0 0 1 0 0 0 1 ...
##  $ state_province.South Dakota: int  0 1 0 0 0 0 0 0 0 0 ...
##  $ state_province.Wyoming    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ num_trans                 : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ basket.no                 : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ basket.yes                : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ refill.no                 : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ refill.yes                : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ area.alcohol              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ area.cooler               : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ area.dispensed            : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ area.fresh                : int  1 0 0 1 1 0 0 0 0 0 ...
##  $ area.fuel                 : int  0 1 0 0 0 0 1 1 0 1 ...
##  $ area.grocery              : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ area.lottery              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ area.miscellaneous        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ area.nongrocery           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ area.snacks               : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ area.tobacco              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ items_sold                : num  1 1 1 1 3 1 1 1 1 1 ...
##  $ loyalty2.not loyal        : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ loyalty2.loyal            : int  0 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, "dummies")=List of 6
##   ..$ quarter       : int [1:4] 3 4 5 6
##   ..$ state_province: int [1:10] 10 11 12 13 14 15 16 17 18 19
##   ..$ basket        : int [1:2] 21 22
##   ..$ refill        : int [1:2] 23 24
##   ..$ area          : int [1:11] 25 26 27 28 29 30 31 32 33 34 ...
##   ..$ loyalty2      : int [1:2] 37 38
```

```
slice_sample(knn_input, n= 10)
```

```
##         high_gpm revenue quarter.1 quarter.2 quarter.3 quarter.4 income
## 116871       low   23.21         0         1         0         0  37917
## 369176       low    5.85         0         0         0         1  91151
## 1161987     high    2.69         0         0         0         1  70557
## 706775      high    2.69         0         0         0         1  45000
## 1007262      low    1.69         0         0         1         0  45705
## 964458       low    2.69         0         0         0         1  37130
## 820507      high    6.29         0         0         1         0  39956
## 427948       low    6.48         0         0         1         0  66250
## 474940      high    2.59         1         0         0         0  64901
## 30196       high    7.00         0         1         0         0  43438
##         bachelors_degree population state_province.Alabama
## 116871                10        546                      0
## 369176              2703      26691                      0
## 1161987             4048      51863                      0
## 706775                55        690                      0
## 1007262              136       1408                      0
## 964458              1238      15153                      0
## 820507               384      11834                      1
## 427948                22        567                      0
## 474940              2883      15839                      0
## 30196                 51        279                      0
##         state_province.Arkansas state_province.Colorado state_province.Iowa
## 116871                        1                       0                   0
## 369176                        0                       0                   0
## 1161987                       0                       1                   0
## 706775                        0                       0                   0
## 1007262                       0                       0                   0
## 964458                        0                       1                   0
```

```
## 820507                         0                         0                    0
## 427948                         0                         0                    1
## 474940                         1                         0                    0
## 30196                          0                         1                    0
##         state_province.Minnesota state_province.Missouri
## 116871                         0                        0
## 369176                         0                        0
## 1161987                        0                        0
## 706775                         0                        1
## 1007262                        0                        0
## 964458                         0                        0
## 820507                         0                        0
## 427948                         0                        0
## 474940                         0                        0
## 30196                          0                        0
##         state_province.Nebraska state_province.Oklahoma
## 116871                        0                       0
## 369176                        0                       0
## 1161987                       0                       0
## 706775                        0                       0
## 1007262                       0                       1
## 964458                        0                       0
## 820507                        0                       0
## 427948                        0                       0
## 474940                        0                       0
## 30196                         0                       0
##         state_province.South Dakota state_province.Wyoming num_trans basket.no
## 116871                            0                      0         1         1
## 369176                            0                      1         1         1
## 1161987                           0                      0         1         1
## 706775                            0                      0         1         1
## 1007262                           0                      0         1         1
## 964458                            0                      0         1         1
## 820507                            0                      0         1         1
## 427948                            0                      0         1         1
## 474940                            0                      0         1         1
## 30196                             0                      0         1         1
##         basket.yes refill.no refill.yes area.alcohol area.cooler area.dispensed
## 116871           0         1          0            0           0              0
## 369176           0         1          0            0           0              0
## 1161987          0         1          0            0           0              0
## 706775           0         1          0            0           0              0
## 1007262          0         1          0            0           1              0
## 964458           0         1          0            0           0              0
## 820507           0         1          0            0           0              0
## 427948           0         1          0            0           0              0
## 474940           0         1          0            0           0              0
## 30196            0         1          0            0           1              0
##         area.fresh area.fuel area.grocery area.lottery area.miscellaneous
## 116871           0         1            0            0                  0
## 369176           0         0            0            0                  0
## 1161987          1         0            0            0                  0
## 706775           0         0            0            0                  0
## 1007262          0         0            0            0                  0
```

```
## 964458           0         0           1           0                  0
## 820507           0         0           0           0                  0
## 427948           0         0           0           0                  0
## 474940           1         0           0           0                  0
## 30196            0         0           0           0                  0
##         area.nongrocery area.snacks area.tobacco items_sold loyalty2.not loyal
## 116871                0           0           0          1                  0
## 369176                0           0           1          1                  0
## 1161987               0           0           0          1                  1
## 706775                0           1           0          1                  1
## 1007262               0           0           0          1                  1
## 964458                0           0           0          1                  1
## 820507                1           0           0          1                  1
## 427948                0           0           1          1                  1
## 474940                0           0           0          2                  0
## 30196                 0           0           0          4                  1
##         loyalty2.loyal
## 116871               1
## 369176               1
## 1161987              0
## 706775               0
## 1007262              0
## 964458               0
## 820507               0
## 427948               0
## 474940               1
## 30196                0
```

```
# Explore the target audience
freq <- table(knn_input$high_gpm)
freq[2]/(freq[1]+freq[2])
```

```
##    high
## 0.43695
```

```
contrasts(knn_input$high_gpm)
```

```
##      high
## low     0
## high    1
```

```
# partition the data
set.seed(77)
partition <- caret::createDataPartition(y =knn_input$high_gpm, p = .75, list = FALSE)
data_train <- knn_input[partition,]
data_test <- knn_input[-partition,]
```

```
#  separate the target variable
x_train <- data_train %>% select(-high_gpm)
```

```
x_test <- data_test %>% select (-high_gpm)
y_train <- data_train$high_gpm
y_test <- data_test$high_gpm
```

```
# z score standardization
x_train <- scale(x_train)
x_test <- scale(x_test)
```

```r
# double check sizes
nrow(x_train)/(nrow(x_test)+nrow(x_train))
```

```
## [1] 0.75005
```

```r
dim(x_train)
```

```
## [1] 15001    37
```

```r
length(y_train)
```

```
## [1] 15001
```

```r
# Run the model
# 141 is the square root of the no. of rows
knn1 = class::knn(train = x_train, test = x_test, cl = y_train, k=141)
```

```r
# Confusion matrix - checking accuracy
confusion_mat = as.matrix(table(knn1, y_test)) #prediction on left and truth on top
print(confusion_mat)
```

```
##       y_test
## knn1    low high
##   low  2448  484
##   high  367 1700
```

```r
# Pre programmed confusion matrix
caret::confusionMatrix(knn1, y_test, positive = "high")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  low high
##       low  2448  484
##       high  367 1700
##
##                Accuracy : 0.8298
##                  95% CI : (0.8191, 0.8401)
##     No Information Rate : 0.5631
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6519
##
##  Mcnemar's Test P-Value : 6.996e-05
##
##             Sensitivity : 0.7784
##             Specificity : 0.8696
##          Pos Pred Value : 0.8224
##          Neg Pred Value : 0.8349
##              Prevalence : 0.4369
##          Detection Rate : 0.3401
##    Detection Prevalence : 0.4135
##       Balanced Accuracy : 0.8240
##
##        'Positive' Class : high
##
```

Overall, our model is really quite accurate. The model makes the correct prediction right here, just about 83 percent of the time. This indicates that there's some assurance that this model would predict well on brand new data.

The model does a great job at predicting when a purchase will be a high profit margin purchase, versus a low profit margin purchase, and it gets it right 83 percent all the time.

```
# Evaluate the data
data_test$prediction <- knn1
data_test <- data_test %>% mutate(correct = high_gpm==prediction)
slice_sample(data_test, n=20)
```

```
##         high_gpm revenue quarter.1 quarter.2 quarter.3 quarter.4 income
## 209931      high    1.79         0         0         1         0  34210
## 1403545      low   26.91         0         0         0         1  29091
## 1226812      low   10.00         1         0         0         0  42772
## 428400       low    5.00         1         0         0         0  66250
## 417907       low   21.00         1         0         0         0  76071
## 38464       high    0.99         0         1         0         0  52500
## 1612610     high    1.00         0         0         0         1  29091
## 1224527      low   45.25         0         0         0         1  70557
## 879815      high    1.98         0         0         1         0  66160
## 318760       low    7.03         0         1         0         0  37130
## 1242817      low   26.90         0         0         0         1  50819
## 382899      high    1.69         0         0         1         0  66250
## 118951       low   13.98         1         0         0         0  41827
## 855741       low    4.75         0         1         0         0  70963
## 1551687     high    2.69         0         0         1         0  66250
## 1201431      low   20.00         0         0         0         1  51461
## 471531      high    7.56         0         0         1         0  63673
## 15110        low    2.13         0         0         0         1  53026
## 1555083     high    1.30         1         0         0         0 132230
## 417764       low   44.64         0         0         1         0  73148
##         bachelors_degree population state_province.Alabama
## 209931              1632      25999                      1
## 1403545                3        181                      0
## 1226812              160       1876                      0
## 428400                22        567                      0
## 417907                93        493                      0
## 38464                 95       1192                      0
## 1612610                3        181                      0
## 1224527             4048      51863                      0
## 879815             10178      57364                      0
## 318760              1238      15153                      0
## 1242817              958       8778                      0
## 382899                22        567                      0
## 118951               169       2379                      0
## 855741              1959      16052                      1
## 1551687               22        567                      0
```

```
## 1201431             2922      52172                     0
## 471531              1172       9046                     0
## 15110                 19        557                     0
## 1555083              974       3294                     0
## 417764               139        852                     0
##         state_province.Arkansas state_province.Colorado state_province.Iowa
## 209931                        0                       0                    0
## 1403545                       0                       0                    0
## 1226812                       0                       0                    0
## 428400                        0                       0                    1
## 417907                        0                       0                    0
## 38464                         0                       0                    0
## 1612610                       0                       0                    0
## 1224527                       0                       1                    0
## 879815                        0                       0                    0
## 318760                        0                       1                    0
## 1242817                       0                       0                    0
## 382899                        0                       0                    1
## 118951                        0                       0                    0
## 855741                        0                       0                    0
## 1551687                       0                       0                    1
## 1201431                       0                       1                    0
## 471531                        0                       0                    1
## 15110                         0                       0                    1
## 1555083                       1                       0                    0
## 417764                        0                       0                    1
##         state_province.Minnesota state_province.Missouri
## 209931                         0                       0
## 1403545                        0                       0
## 1226812                        0                       0
## 428400                         0                       0
## 417907                         0                       0
## 38464                          0                       0
## 1612610                        0                       0
## 1224527                        0                       0
## 879815                         0                       1
## 318760                         0                       0
## 1242817                        1                       0
## 382899                         0                       0
## 118951                         0                       0
## 855741                         0                       0
## 1551687                        0                       0
## 1201431                        0                       0
## 471531                         0                       0
## 15110                          0                       0
## 1555083                        0                       0
## 417764                         0                       0
##         state_province.Nebraska state_province.Oklahoma
## 209931                        0                       0
## 1403545                       0                       1
## 1226812                       0                       1
## 428400                        0                       0
## 417907                        1                       0
## 38464                         1                       0
```

```
## 1612610                       0                       1
## 1224527                       0                       0
## 879815                        0                       0
## 318760                        0                       0
## 1242817                       0                       0
## 382899                        0                       0
## 118951                        0                       1
## 855741                        0                       0
## 1551687                       0                       0
## 1201431                       0                       0
## 471531                        0                       0
## 15110                         0                       0
## 1555083                       0                       0
## 417764                        0                       0
##         state_province.South Dakota state_province.Wyoming num_trans basket.no
## 209931                            0                       0         1         1
## 1403545                           0                       0         1         1
## 1226812                           0                       0         1         1
## 428400                            0                       0         1         1
## 417907                            0                       0         1         1
## 38464                             0                       0         1         1
## 1612610                           0                       0         1         1
## 1224527                           0                       0         1         1
## 879815                            0                       0         1         1
## 318760                            0                       0         1         1
## 1242817                           0                       0         1         1
## 382899                            0                       0         1         1
## 118951                            0                       0         1         1
## 855741                            0                       0         1         1
## 1551687                           0                       0         1         1
## 1201431                           0                       0         1         1
## 471531                            0                       0         1         1
## 15110                             0                       0         1         1
## 1555083                           0                       0         1         1
## 417764                            0                       0         1         1
##         basket.yes refill.no refill.yes area.alcohol area.cooler area.dispensed
## 209931           0         1          0            0           0               0
## 1403545          0         1          0            0           0               0
## 1226812          0         1          0            0           0               0
## 428400           0         1          0            0           0               0
## 417907           0         1          0            0           0               0
## 38464            0         1          0            0           1               0
## 1612610          0         1          0            0           0               1
## 1224527          0         1          0            0           0               0
## 879815           0         1          0            0           0               0
## 318760           0         1          0            0           0               0
## 1242817          0         1          0            0           0               0
## 382899           0         1          0            0           0               1
## 118951           0         1          0            0           0               0
## 855741           0         1          0            0           0               0
## 1551687          0         1          0            0           0               0
## 1201431          0         1          0            0           0               0
## 471531           0         1          0            0           1               0
## 15110            0         1          0            0           1               0
```

```
## 1555083          0         1          0            0          0             0
## 417764           0         1          0            0          0             0
##         area.fresh area.fuel area.grocery area.lottery area.miscellaneous
## 209931           1         0            0            0                  0
## 1403545          0         1            0            0                  0
## 1226812          0         1            0            0                  0
## 428400           0         1            0            0                  0
## 417907           0         1            0            0                  0
## 38464            0         0            0            0                  0
## 1612610          0         0            0            0                  0
## 1224527          0         1            0            0                  0
## 879815           0         0            0            0                  0
## 318760           0         0            0            0                  0
## 1242817          0         1            0            0                  0
## 382899           0         0            0            0                  0
## 118951           0         0            0            0                  0
## 855741           0         0            0            0                  0
## 1551687          1         0            0            0                  0
## 1201431          0         1            0            0                  0
## 471531           0         0            0            0                  0
## 15110            0         0            0            0                  0
## 1555083          1         0            0            0                  0
## 417764           0         1            0            0                  0
##         area.nongrocery area.snacks area.tobacco items_sold loyalty2.not loyal
## 209931                0           0            0          1                  1
## 1403545               0           0            0          1                  1
## 1226812               0           0            0          1                  1
## 428400                0           0            0          1                  1
## 417907                0           0            0          1                  0
## 38464                 0           0            0          1                  0
## 1612610               0           0            0          1                  1
## 1224527               0           0            0          1                  1
## 879815                0           1            0          2                  1
## 318760                0           0            1          1                  0
## 1242817               0           0            0          1                  1
## 382899                0           0            0          1                  0
## 118951                0           0            1          2                  1
## 855741                0           0            1          1                  1
## 1551687               0           0            0          2                  1
## 1201431               0           0            0          1                  1
## 471531                0           0            0          4                  1
## 15110                 0           0            0          1                  0
## 1555083               0           0            0          1                  1
## 417764                0           0            0          1                  0
##         loyalty2.loyal prediction correct
## 209931               0       high    TRUE
## 1403545              0        low    TRUE
## 1226812              0        low    TRUE
## 428400               0        low    TRUE
## 417907               1        low    TRUE
## 38464                1        low   FALSE
## 1612610              0       high    TRUE
## 1224527              0        low    TRUE
## 879815               0       high    TRUE
```

```
## 318760              1        low    TRUE
## 1242817             0        low    TRUE
## 382899              1       high    TRUE
## 118951              0        low    TRUE
## 855741              0        low    TRUE
## 1551687             0       high    TRUE
## 1201431             0        low    TRUE
## 471531              0        low   FALSE
## 15110               1        low    TRUE
## 1555083             0       high    TRUE
## 417764              1        low    TRUE
```

The first line of code adds the predicted low or high gross profit margin back into the test data. The second line, creates a new variable called correct, that takes on the value true, when the model was correct and false, when the model was incorrect. Now printing on a sample of this dataframe and scrolling through it gives us an idea or the ability to indicate when the KNN model was right and when it was wrong.

We can look at the first column is the truth, so high_gpm, that's the truth, and if we scroll all the way over to the right, we can see the prediction right here from KNN, that's predicted on our test data. Then we can look at whether it's a correct or not prediction.

Promoting certain areas, such as fresh and dispensed, might help increase the sale of high profit margins. While other areas such as lottery, might actually hurt. However, our efforts here to make a prediction model, to examine the relevance of the factors, we would need to employ another method such as logistic regression or decision trees