ROSSMANN DRUGSTORE SALES
&
POSITIONING PIZZA BRANDS

MKT 372 Data-Driven Marketing

The Final Research Report

Erica Lowery

Elisa Ma

Fall 2019

Table of Contents

SECTION I.

Background & Problem

Throughout the business world, there are multiple problems companies are trying to solve in order to enhance overall performance. A few examples include, “how to optimize business sales and revenue,” “how to decrease customer churn and increase retention,” and many others. Within this final research report, we will be analyzing and discussing two different datasets and the data-driven marketing problem for each:

1. *How can we increase Rossmann stores’ performance?*
2. *How can we position different pizza brands based on competitors?*

Rossmann Stores

Starting off with our exploration regarding Rossmann stores dominant in Europe, the data was collected on *Github* where we examined each variable provided. There are a total of 12 variable columns and over 1,100 observation rows, with the dependent variable of *Sum of Total Sales*, and each row pertaining to a different Rossmann store. We will be explaining the independent variables more in depth in the following section. On brief inspection, it is evident to see some Rossmann stores perform better than others. This leads to the potential research objectives we will be addressing for the Rossmann dataset:

1. *Can we predict store sales? Which variables have a significant relationship on Rossmann store sales?*
2. *What are the most influential variables when measuring store sales and what are the potential explanations for why?*
3. *Do certain variables have high correlations that we should be aware of?*

Corporations are constantly trying to formulate strategies to lower costs or increase profits, or both. Thus, this research is critical because firms desire to know why and what variables contribute to certain stores performing better than others. If they are able to figure out which factors are most positively significant to revenue, businesses can better allocate their money and resources to those variables for a more optimized marketing approach. On the other hand, if they can determine which factors are insignificant or contribute negatively to sales, firms will know not to wastefully invest their funds and resources into those variables. To address the research problem of why certain Rossmann stores perform better than others, some quantitative analyses that we will be performing include multiple linear regressions, backward regression, and AIC and BIC to compare our models. Overall, we will be looking at variables of Rossmann stores and their sales performance to formulate recommendations on how Rossmann can improve its sales of lower-performing stores based on higher-performing stores. Additionally, this research will also provide insight on various marketing decisions, such as, "what variable(s) and key component(s) should Rossmann emphasize in its marketing campaigns in order to attract more customers and increase sales to its stores?"

Pizza Brands

For our pizza research problem, the dataset was obtained on *data.world* with a total of nine variable columns and 300 data points. The dependent variable is the pizza *brand* with a total of ten different brands (A, B, C, D, E, F, G, H, I, J), with each row pertaining to a different pizza ID. We will be explaining the independent variables more in depth in the following section. Our objectives for this research include:

1. *Which variables are associated with which pizza brands?*
2. *Which brands are relevant competitors based on similar characteristics?*
3. *What areas are open for opportunity of potential new brands to enter?*

In a saturated market with almost every industry filled with multiple competitors, it is essential for companies to understand how customers perceive them. Therefore, this research is crucial because it helps pizza brands learn about their brand perception and where they stand compared to their competitors. For example, if a certain pizza brand wants to be classified as a low carbohydrate alternative to other brands, is it successfully reaching its goal and appealing to its target market? Especially as trends, such as healthy and conscious eating, are becoming increasingly prevalent, it is important for pizza and other food brands to be aware of how consumers perceive them. A few quantitative analysis methods we will be implementing multivariate data analysis, such as principal component analysis (PCA). This data analysis will provide business leads on how to improve marketing strategies for pizza brands. For example, what other brands do audience segments associate your pizza brand with? How should a pizza brand competitively differentiate themself?

SECTION II.

Data Summary & Exploratory Analysis

Rossmann Stores

As mentioned in the previous section, we obtained our Rossmann dataset from GitHub. Hence, it is secondary data. In total, there are 12 variables:

Variable	**Description**
Store ID	Unique ID of Rossmann stores
Sum of Total Sales	Total sales ($) of the store
Sum of # of Customers	Total number of customers of the store
Sum of # of Days open	Total number of days store open
Sum of # of Promotion days	Total number of days store ran promotion – 286 (reference) or 360
Sum of # of State Holidays	Total number of days store closed due to state holidays
Sum of # of School Holidays	Total number of days store affected by public school holiday (* Public schools closed on state holidays and weekends)
StoreType	Specifies which of 4 different store models – A (reference), B, C, D
Assortment	Specifies which assortment level – A = basic (reference), B = extra, C = extended
CompetitionDistance	Distance (meters) to the nearest drugstore competitor
MonthsSinceCompetition	Months since the nearest competitor has opened
Promo2	Continuing promotion for the store – 0 = No/Not Participating (reference), 1 = Yes/Participating

Exhibit 1: Descriptions of Rossmann Variables

Before beginning to analyze our data, we transformed *Store ID*, *Sum of # of Promotion Days*, *StoreType*, *Assortment*, and *Promo2* into categorical variables, leaving the remaining variables to be numerical.

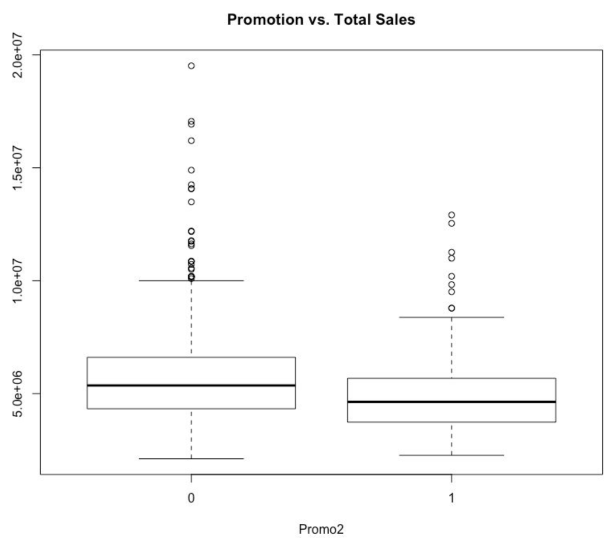

Exhibit 2: Rossmann Boxplot

Looking at the boxplot (*Exhibit 2*) of *Total Sales* grouped by Rossmann stores that had continued promotions (*Promo2 = 1*) and stores that did not (*Promo2 = 0*), we do not see a compelling difference in the average sales between the two groups. In fact, we see a slightly higher performance of *Total Sales* for stores not continuously running the promotions, which may not be what we originally would have expected. The stores that did not continuously run the promotions, there also appears to be a higher spread of *Total Sales*, specifically on the upper end of sales above the mean.

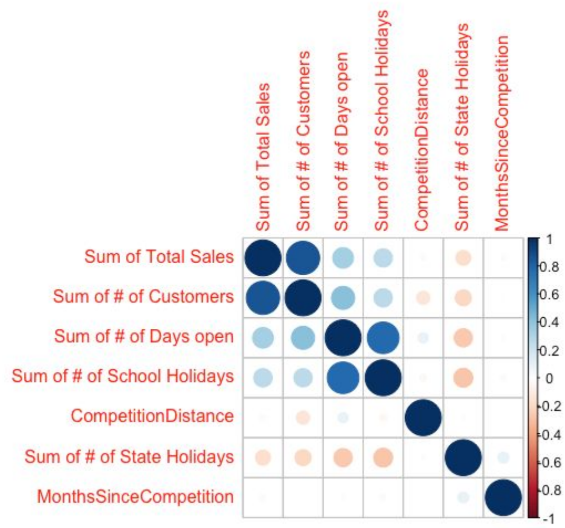

Exhibit 3: Rossmann Correlation Matrix Plot

For the most part, the correlation matrix plot (*Exhibit 3*) shows there is not a high correlation between most of the variables, with the exceptions of *Total Sales* being highly positively correlated with *Customers* and *Sum of School Holidays* being highly positively correlated with the *Sum of State Holidays*. The high, positive correlation (0.854) between the total number of sales and number of customers naturally makes sense considering each additional customer brings in at least some additional revenue. Therefore, Rossmann stores with more customers tend to have higher sum of total sales. This is a critical piece of information to note as we move forward with our data analysis.

Pizza Brands

The pizza dataset from data.world. Hence, it is secondary data. In total, there are 9 variables:

Variable	Description
brand	Specifies which of 10 pizza brands – A, B, C, D, E, F, G, H, I, J
id	Sample ID of pizza
mois	Water measured per 100 gram sample
prot	Protein measured per 100 gram sample
fat	Fat measured per 100 gram sample

ash	Ash measured per 100 gram sample
sodium	Sodium measured per 100 gram sample
carb	Carbohydrates measured per 100 gram sample
cal	Calories measured per 100 gram sample

Exhibit 4: Descriptions of Pizza Variables

In order to further diagnose and explore relationships between our variables and the summary statistics of our dataset, we have performed and attached various tables and graphs that depict information about the pizza dataset's variables.

Brand	**Moisture**	**Protein**	**Fat**	**Ash**	**Sodium**	**Carb**	**Calories**
A	29.966	20.107	43.447	5.014	1.656	1.487	4.774
B	51.308	13.639	27.620	3.464	0.984	3.970	3.191
C	49.477	26.026	19.171	3.283	0.465	2.046	2.849
D	47.671	22.231	21.645	4.316	0.715	4.136	3.003
E	36.083	7.733	15.116	1.476	0.449	39.592	3.254
F	29.404	7.898	16.425	1.474	0.462	44.787	3.596
G	28.241	8.237	15.644	1.447	0.444	46.432	3.595
H	35.825	7.895	14.292	1.406	0.416	40.584	3.225
I	54.593	10.383	13.061	2.098	0.487	19.866	2.384
J	46.035	10.566	16.324	2.365	0.614	24.736	2.878

Exhibit 5: Pizza Mean of Measurements

The table (*Exhibit 5*) shows the mean of measured variables separated by the 10 brands. Brand A holds the highest mean (blue) for most of the variables, while having the lowest (red) measured carbohydrates per sample compared to the other brands. As a result, it becomes difficult to interpret where brand A is positioned when it has relatively high measurements of fat, ash, sodium, and calories compared to other brands. We can see brands like B and F have neither the highest or lowest means for measured characteristics and lie somewhere in between the other brands. Using this information, we can start to see where certain brands may be positioned based on which variables they are measured “higher” and “lower” in relative to other pizza brands, and how to differentiate each of the brands.

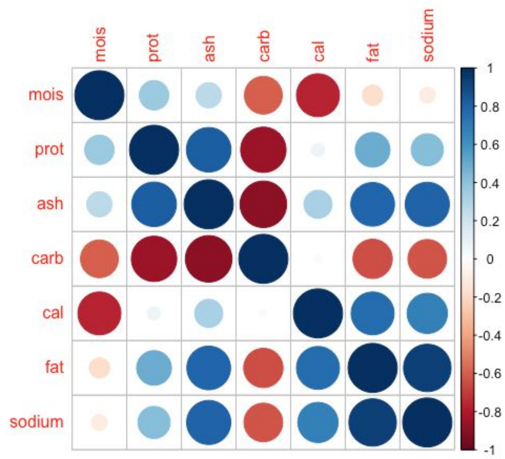

Exhibit 6: Pizza Correlation Matrix Plot

The correlation matrix plot of the different characteristics (*Exhibit 6*) shows us which variables are positively correlated (blue) and negatively correlated (red). For example, high carbohydrate measured in pizza samples is highly negatively correlated with the measure of protein. In other words, pizza samples measured with higher carb per 100 grams typically contained a lower measure of protein per sample. On the other hand, variables such as fat and sodium are highly positively correlated, where high measures of fat in pizza samples typically have higher amounts of sodium as well. Considering the high correlation, positive and negative, of many of the variables, principal component analysis is a practical method to analyze the data of the pizza brands.

SECTION III.

Data Analysis, Key Findings, & Conclusions

Rossmann Stores

The first draft version of our linear model contained all variables in order to predict *Sum of Total Sales* resulted in relatively high AIC (3345.77) and BIC (33511.02), indicating a problem with our raw model. By plotting the first model, we noticed it did not meet the assumptions for an accurate linear regression model. The points on the Residuals vs Fitted (*Exhibit 7.1*, left) plot are clustered together to the left, with a slightly downward trend, indicating the function is not linear (nonlinearity). The points on the Normal Q-Q plot (*Exhibit 7.1*, right) somewhat lie along the diagonal line until the upper portion of the graph where the points begin to slope upwards above the line, indicating the function is not normally distributed.

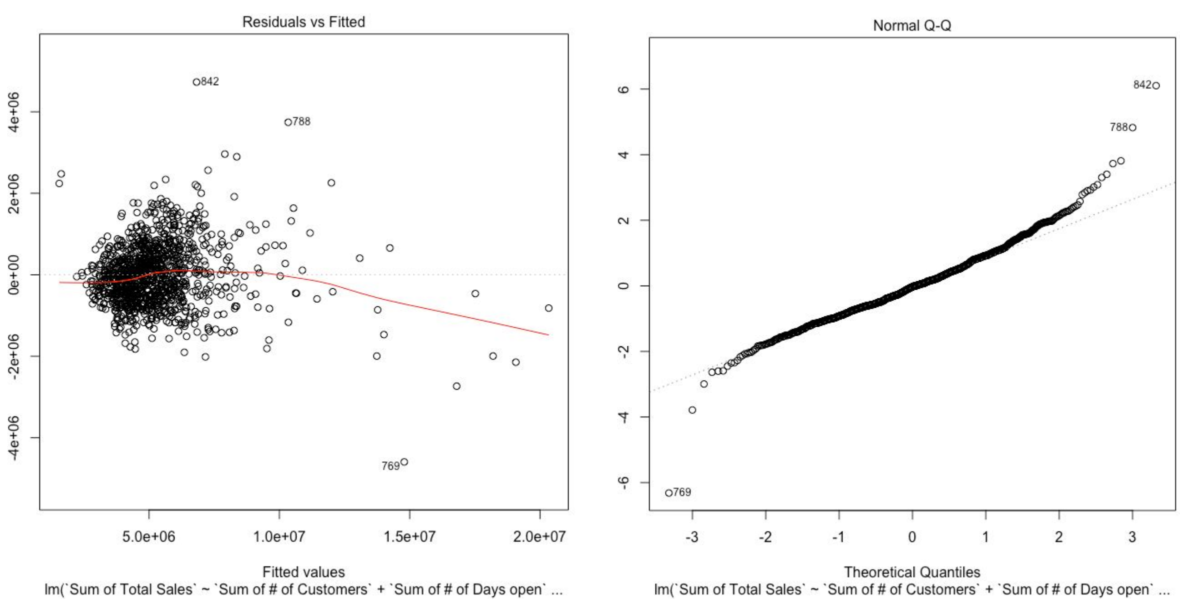

Exhibit 7.1: Rossmann Residuals vs Fitted Plot & Normal Q-Q Plot

To address nonlinearity and the right-skewed histogram distribution of a few variables, we transformed four variables and took the log of *Sum of Total Sales*, *Sum of # of Customers*, *CompetitionDistance*, and *MonthsSinceCompetition* in order to make the distribution more normal to fit the assumptions of a linear regression. With the second model containing the transformed variables and all remaining variables, we are able to satisfy the assumptions of linearity and normality. Looking at the Residuals vs Fitted plot (*Exhibit 7.2*, left), we see no pattern and a relatively equal spread of all the points, signaling the function is linear and the assumption is met. In addition, the plots fall on the line of the Normal Q-Q plot (*Exhibit 7.2*, right), indicating the normality assumption is also met.

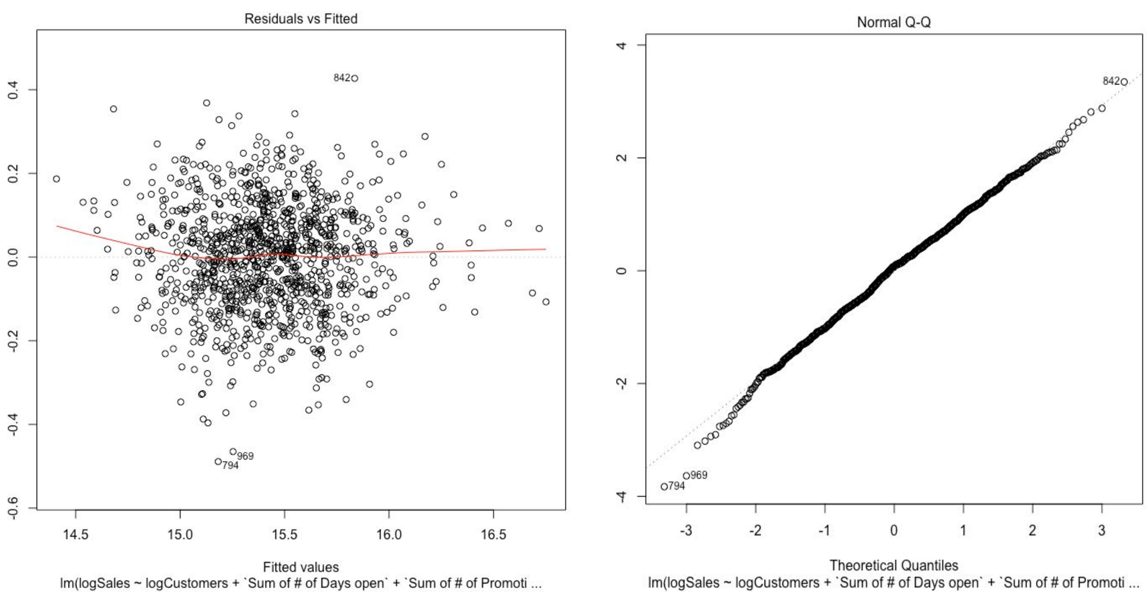

Exhibit 7.2: Rossmann Residuals vs Fitted Plot & Normal Q-Q Plot

With the transformations and assumptions being met, we noticed a significant drop in our AIC (-1392.26) and BIC (-1317.011) values, indicating a much better model overall. Taking our second model a step further, we ran a backwards stepwise regression in order to simplify to a parsimonious model based on significant variables and AIC. Through this method, we finalized our third model with all significant variables ($p < 0.05$), with an overall lower AIC (-1395.11) and BIC (-1329.895).

(Predicted) logSales = ***3.171*** *+* ***0.878***×*logCustomers –* ***0.117***×*PromotionDays360 +* ***0.006***×*StateHolidays + * ***0.002***×*SchoolHolidays –* ***0.146***×*StoreTypeB –* ***0.022***×*StoreTypeC +* ***0.160***×*StoreTypeD –* ***0.438***×*AssortmentB +* ***0.050***×*AssortmentC +* ***0.033***×*logCompDistance +* ***0.046***×*Promo2Yes*

The linear regression is our version of the best fit model to predict sales (*logSales*) and identify which significant variables positively and negatively impact stores' performance. Our interpretation of the estimated coefficients for each variable of our model:

- *Intercept*: When all other variables are zero, the base level of *logSales* is a positive 3.171. In practice, this coefficient is actually meaningless because it is not realistic for all other variables to be held constant at zero.
- *logCustomers*: As the number of Rossmann store customers increases, the total sales are estimated to increase. More specifically, each increase in *logCustomers* increases *logSales* by 0.878.
- *PromotionDays360*: The overall sales of individual Rossmann stores is estimated to decrease when the store runs the promotion for 360 days (*PromotionDays* = 360), compared to only running the promotion for 286 days (reference, *PromotionDays* = 286).
- *StateHolidays*: As the number of state holidays increases, the total sales of individual Rossmann stores are expected to increase slightly. More specifically, each additional day of *StateHolidays* increases *logSales* by 0.006.
- *SchoolHolidays*: As the number of school holidays increases, the total sales of individual Rossmann stores are expected to increase slightly (less than that of state holidays). More specifically, each additional day of *SchoolHolidays* increases logSales by 0.002.
- *StoreTypeB*: When the Rossmann store is of type B, total sales are estimated to be less when compared to stores of type A (reference).
- *StoreTypeC*: When the Rossmann store is of type C, total sales are estimated to be less when compared to stores of type A (reference); however, stores of type C (-0.022) have less of a negative impact on sales (*logSales)* as store types B (-0.146).
- *StoreTypeD*: When the Rossmann store is of type D, total sales are estimated to be more compared to stores of type A (reference). Overall, store types D have

the most “positive” impact on total sales compared all the four Rossmann store types.

- *AssortmentB*: Rossmann stores of extra assortment level (*Assortment* = B), total sales are estimated to be more when compared to basic assortment level stores (reference, *Assortment* = A).
- *AssortmentC*: Rossmann stores of extended assortment level (*Assortment* = C), total sales are estimated to be more when compared to basic assortment level stores (reference, *Assortment* = A); however, extra assortment level (*Assortment* = B) has the most “positive” impact on sales compared to all three of the Rossmann assortment levels.
- *logCompDistance*: The farther a Rossmann competitor, the higher the estimated total sales of the Rossmann store. More specifically, each increase in *logCompDistance* increases *logSales* by 0.033.
- *Promo2Yes*: Rossmann stores participating in the continued promotion (*Promo2* = 1) are expected to have higher sales compared to stores not participating in the continued promotion (reference, *Promo2* = 0). More specifically, *logSales* is expected to increase by 0.046 with participating stores compared to non participating stores.

Pizza Brands

Before conducting principal component analysis, the relevant variables had to be rescaled (excluding *brand* and *id*) in order to summarize the values measured via dimension reduction of the multivariate data. Through principal component analysis with those variables we scaled, we found combinations of common variation and rotated the original variables in order to plot the first two principal components and the different variables onto a 2D positioning map; however, the plot below mapping the pizza samples individually (*Exhibit 8*, left) is difficult to interpret according to their brands. By finding the mean “rating” of each pizza sample based on the measured characteristics and grouping them by their brands, the pizza brand positioning map plotted by brands (*Exhibit 8*, right) becomes easier to interpret and differentiate the individual brands.

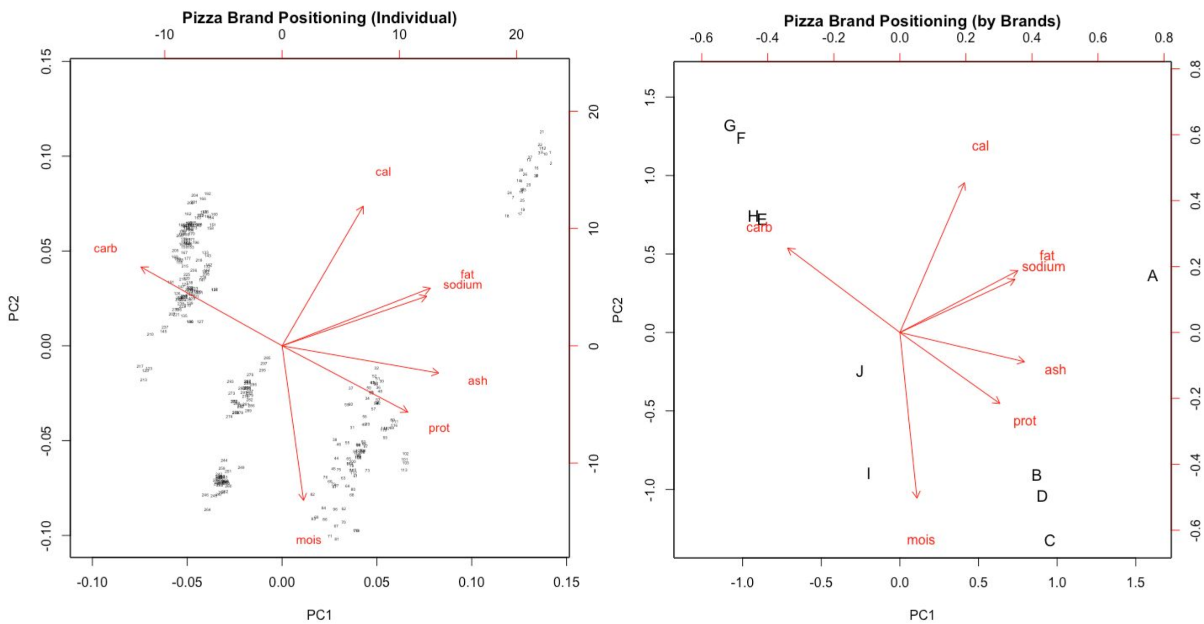

Exhibit 8: Pizza Principal Component Analysis (PCA)

Our principal component analysis visually maps out the relationship between different variables and the different pizza brands associated with each. Similar variables with positive relationships, such as the amounts of fat and sodium measured, are grouped closer together on the map. On the other hand, variables with negative relationships, such as amounts of carbohydrates and protein, are mapped opposite of each other. Through analyzing combinations of the measured variables, we are able to map the different characteristics taken on by different brands. Additionally, the brand positioning map also shows which brands are relevant competitors of each other based on similar characteristics dominant in brands, such as brands E and H measured as high carb pizzas. We can also see the lack of much brand presence of pizzas with high fat and sodium, which would logically be unappealing to consumers. Using this brand positioning map, we can clearly see how similar and different each of the pizza brands are, which brands don't have relevant competitors, and areas where there are no existing pizza brands, which will help formulate marketing recommendations based on our findings using data.

SECTION IV.

Marketing Strategy Recommendations

Now that we have analyzed our data and inspected key findings, we will be synthesizing our results into data-driven, actionable suggestions and next steps into our marketing strategies.

Rossmann Stores

In Section III, we identified the linear regression model to best explain the relationship each variable has with the dependent variable, *logSales*. Moving forward, we suggest Rossmann to emphasize the independent variables that contribute positively towards *logSales* in the company's marketing tactics, such as participating in the continued and consecutive promotions (*Promo2* = 1) to increase sales as opposed to not participating in continued promotions. Considering stores with more *StateHolidays* and *SchoolHolidays* in the area have positive impacts on total sales (*logSales*), Rossmann stores with less holidays should find other strategies to make up for the lower performance in sales, possibly through running additional promotions. Considering Rossmann store types D and extended assortment levels (*Assortment* = C) historically have higher performance in terms of total sales, other stores and assortment types should consider adjusting if possible and logical depending on the individual stores. On the other hand, Rossmann stores running promotions all year long (*Sum of # of Promotion days* = 360) actually has a negative impact on sales, therefore stores should consider running less sales more strategically 286 days instead in order to increase total sales for individual stores. Rossmann should also stay aware of closer competitors (*logCompDist*) which have a negative impact on sales, and find leverage in order to keep sales of stores with close competitors still competitively high. New Rossmann stores entering the market should also consider strategically opening where there are farther competitors and areas with more holidays, as well as open as store model type D with an extended assortment level (*Assortment* = C).

Pizza Brands

The Principal Component Analysis (PCA) map illustrated which brands are grouped together and by which common characteristics (carbohydrates, proteins, etc.). Thus, brands can visually see similar competitors and use that information to their advantage in order to leverage themselves above relevant competitors. Our recommendation for brands that are strongly related to another, such as brand E and H, is to find ways to distinguish themselves from each other in their marketing campaigns. If consumers are unable to distinguish between two separate brands and view them as entities that ultimately provide the same products, there will likely be low customer retention and brand loyalty. For brands that are in the middle of the PCA map, such as brand J, we can infer that they are a mixture of all the variables. This kind of brand perception may be unfavorable because consumers may not understand what the brand's main selling point or differentiating factor is, unless brand J is purposely positioning itself that way. However, being known as a brand that offers everything also comes with the risk of being a brand that does not excel in anything. Hence, we suggest these brands decide on a specific variable to be classified as instead of being muddled in the middle, or find a different value proposition other than the pizza characteristics itself in order to provide resonance with customers. This will provide the marketing departments with more direction on how to classify and market themselves to the public. Overall, in order to know what unique advantage to highlight for each brand in its marketing assets, we need to perform additional research.

Limitations & Future Research Directions

Although through our own research and examination we have been able to make useful, data-driven conclusions and recommendations, there are still many avenues that can be explored further in depth; however, not all proposals and strategies are feasible with the given information. Whether it be due to a lack of resources, time, or other reasons, there are realistic limitations to what we were able to analyze and explore.

Rossmann Stores

It is crucial to note the insufficient specific details on what each store type corresponds to. For example, *StoreTypeD* is the best alternative, yet we do not have further information on what that encompasses. It may be worthwhile to conduct further research to discover the key differences between each store type and why each store model performs differently in terms of total sales. Additionally, the final linear model regression is for the overall Rossmann brand. However, each individual Rossmann store is in different circumstances, whether it be distance from closest competitor, number of school holidays, or others. Due to this, it may be difficult to generalize what an entire brand chain should do depending on specific variables. Therefore, it may be beneficial to divide the Rossmann stores into groups based off of store type, assortment level, or by another factor. Nonetheless, there is an uneven distribution of store types and assortment levels found in the data which would skew results of exploring different variables and their relationship to total store sales.

Pizza Brands

Limitations to our pizza brand positioning map include variables uncorrelated to each other not represented accurately on the map. According to the correlation matrix plot of the different variables, the amount of carbohydrates and calories found in pizza samples are not very well correlated (-0.023); however, our brand positioning map shows calories as the closest variable line to carbohydrates. This is highly misleading, suggesting that carbohydrates are most correlated with the amount of calories. Referring back to our correlation plot, carbohydrates are actually highly negatively correlated with other variables (other than calories), which explains why it appears closest to the calorie variable. Our potential next steps include pinpointing what specific brands corresponds with each letter. For example, is “Brand A” Domino’s Pizza or Cici’s Pizza, or a local pizza store? If we are able to gather more information on who we are catering to and what each brand’s performance is, we would be able to provide stronger and more accurate marketing recommendations. In addition, it would be valuable for each pizza brand to perform a SWOT analysis (strengths, weaknesses, opportunities, and threats) to help each pizza brand understand their own competitive advantages and how they compare to their respective competitors. The pizza industry is heavily saturated with many businesses entering and competing, especially with quick changes in market trends and consumer preferences. Thus, in this kind of scenario, it is necessary that brands have a clear understanding of their own brand and their competitors in order to produce marketing plans and strategies to stay relevant in the market and superior to their relevant competitors.

Appendices

Rossmann Stores

```
Start:  AIC=-4558.49
logSales ~ logCustomers + `Sum of # of Days open` + `Sum of # of Promotion days` +
    `Sum of # of State Holidays` + `Sum of # of School Holidays` +
    StoreType + Assortment + logCompDist + logMonthsComp + Promo2

                                Df Sum of Sq     RSS     AIC
- `Sum of # of Days open`        1     0.001  18.233 -4560.4
- logMonthsComp                  1     0.018  18.249 -4559.4
<none>                                        18.232 -4558.5
- `Sum of # of Promotion days`   1     0.084  18.316 -4555.4
- `Sum of # of State Holidays`   1     0.193  18.425 -4548.8
- `Sum of # of School Holidays`  1     0.363  18.595 -4538.5
- Promo2                         1     0.479  18.711 -4531.6
- Assortment                     2     1.420  19.652 -4478.9
- logCompDist                    1     2.574  20.806 -4413.2
- StoreType                      3     5.308  23.539 -4279.6
- logCustomers                   1    81.961 100.193 -2660.6

Step:  AIC=-4560.43
logSales ~ logCustomers + `Sum of # of Promotion days` + `Sum of # of State Holidays` +
    `Sum of # of School Holidays` + StoreType + Assortment +
    logCompDist + logMonthsComp + Promo2

                                Df Sum of Sq     RSS     AIC
- logMonthsComp                  1     0.018  18.251 -4561.3
<none>                                        18.233 -4560.4
- `Sum of # of State Holidays`   1     0.215  18.448 -4549.4
- `Sum of # of School Holidays`  1     0.372  18.605 -4539.9
- Promo2                         1     0.478  18.711 -4533.6
- `Sum of # of Promotion days`   1     0.571  18.803 -4528.1
- Assortment                     2     1.424  19.657 -4480.6
- logCompDist                    1     2.606  20.839 -4413.5
- StoreType                      3     5.335  23.567 -4280.3
- logCustomers                   1    82.826 101.059 -2653.0

Step:  AIC=-4561.34
logSales ~ logCustomers + `Sum of # of Promotion days` + `Sum of # of State Holidays` +
    `Sum of # of School Holidays` + StoreType + Assortment +
    logCompDist + Promo2

                                Df Sum of Sq     RSS     AIC
<none>                                        18.251 -4561.3
- `Sum of # of State Holidays`   1     0.206  18.457 -4550.8
- `Sum of # of School Holidays`  1     0.359  18.610 -4541.6
- Promo2                         1     0.471  18.721 -4534.9
- `Sum of # of Promotion days`   1     0.560  18.810 -4529.7
- Assortment                     2     1.410  19.661 -4482.3
- logCompDist                    1     2.626  20.877 -4413.5
- StoreType                      3     5.317  23.568 -4282.3
- logCustomers                   1    82.808 101.059 -2655.0
```

Appendix A

```
Call:
lm(formula = logSales ~ logCustomers + `Sum of # of Promotion days` +
    `Sum of # of State Holidays` + `Sum of # of School Holidays` +
    StoreType + Assortment + logCompDist + Promo2, data = RM)

Residuals:
     Min       1Q   Median       3Q      Max
-0.48498 -0.08534  0.00796  0.08435  0.43162

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      3.1707391  0.1861826  17.030  < 2e-16 ***
logCustomers                     0.8778465  0.0124088  70.744  < 2e-16 ***
`Sum of # of Promotion days`360 -0.1165084  0.0200287  -5.817 7.84e-09 ***
`Sum of # of State Holidays`     0.0060071  0.0017023   3.529 0.000435 ***
`Sum of # of School Holidays`    0.0016663  0.0003575   4.661 3.53e-06 ***
StoreTypeb                      -0.1464777  0.0474803  -3.085 0.002086 **
StoreTypec                      -0.0219519  0.0118629  -1.850 0.064513 .
StoreTyped                       0.1599569  0.0095975  16.667  < 2e-16 ***
Assortmentb                     -0.4375075  0.0629619  -6.949 6.29e-12 ***
Assortmentc                      0.0500398  0.0083313   6.006 2.58e-09 ***
logCompDist                      0.0332996  0.0026433  12.598  < 2e-16 ***
Promo21                          0.0464955  0.0087163   5.334 1.16e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1286 on 1103 degrees of freedom
Multiple R-squared:  0.8528,    Adjusted R-squared:  0.8513
F-statistic:   581 on 11 and 1103 DF,  p-value: < 2.2e-16
```

Appendix B

Pizza Brands

```
pizza$brand: A
     mois      prot       fat       ash    sodium      carb       cal
29.966207 20.107241 43.446897  5.014483  1.656207  1.486897  4.773793
------------------------------------------------------------------------------------------
pizza$brand: B
      mois       prot        fat        ash     sodium       carb        cal
51.3077419 13.6387097 27.6203226  3.4635484  0.9848387  3.9696774  3.1909677
------------------------------------------------------------------------------------------
pizza$brand: C
      mois       prot        fat        ash     sodium       carb        cal
49.4774074 26.0255556 19.1711111  3.2833333  0.4648148  2.0462963  2.8488889
------------------------------------------------------------------------------------------
pizza$brand: D
     mois      prot       fat       ash    sodium      carb       cal
47.671250 22.231250 21.645312  4.315937  0.715000  4.136250  3.003437
------------------------------------------------------------------------------------------
pizza$brand: E
      mois       prot        fat        ash     sodium       carb        cal
36.0832143  7.7328571 15.1157143  1.4760714  0.4492857 39.5921429  3.2539286
------------------------------------------------------------------------------------------
pizza$brand: F
     mois      prot       fat       ash    sodium      carb       cal
29.404333  7.898000 16.424667  1.473667  0.462000 44.787333  3.596000
------------------------------------------------------------------------------------------
pizza$brand: G
      mois       prot        fat        ash     sodium       carb        cal
28.2410345  8.2365517 15.6437931  1.4468966  0.4437931 46.4317241  3.5951724
------------------------------------------------------------------------------------------
pizza$brand: H
      mois       prot        fat        ash     sodium       carb        cal
35.8251515  7.8945455 14.2915152  1.4060606  0.4160606 40.5839394  3.2245455
------------------------------------------------------------------------------------------
pizza$brand: I
      mois       prot        fat        ash     sodium       carb        cal
54.5927586 10.3831034 13.0606897  2.0982759  0.4872414 19.8655172  2.3841379
------------------------------------------------------------------------------------------
pizza$brand: J
     mois      prot       fat       ash    sodium      carb       cal
46.035000 10.566250 16.324063  2.364688  0.614375 24.735937  2.878437
```

Appendix C