inMarket Data Challenge – Business Insights

Yongbock (David) Kwon
Nov 17th, 2018

Description

"A fictional children's electrolyte drink company, named Brawndo, is interested in learning insights about their customers and would like you to present this information to their CMO. Table 1 shows the behavior of two distinct sets of consumers. Namely, a sample set of consumers who buy the drink, as well as a sample set of all US consumers generally. Table 1 shows places each set consumers are seen at, not necessarily where they purchased the product. Keep in mind each person can visit multiple places. Table 2 shares the ages of users. Using the two tables, come up with 3 insights about how Brawndo's consumers and how they should target future marketing based on the data."

Assumption

To clarify and to avoid confusion of the understanding of the dataset, I developed some assumptions as following:

1. Variable assumption

In Table1 and Table2, I assume that,

"Non-Customers" variable is the number of customers who don't buy drink.

"Customers" variable is the number of customers who buy drink.

2. Assumptions from the business

Since the client company, Brawndo, is the "children's electrolyte drink company," I assume that the future target customer is the children, which is the customers in age range 0-18. Also, another assumption in here is that this dataset doesn't depend on the fact that parents don't buy drink for their children.

Methodology

I started to explore the dataset, Table 1, by creating two summary tables. The summary tables shows the each occurrence and the average of number of customers who buy drink or who don't buy drink in chain categories.

Summary for Customers by Chain Category		
Chain Category	Count	Mean
Department Stores	11	476
Drinking Places (alcoholic Beverages)	1	58
Drug Stores and Proprietary Stores	10	1151
Eating Places	194	517
Grocery Stores	45	244
Hardware Stores	7	507

Summary for Non-Customers by Chain Category		
Chain Category	Count	Mean
Department Stores	11	581
Drinking Places (alcoholic Beverages)	1	105
Drug Stores and Proprietary Stores	10	1120
Eating Places	194	488
Grocery Stores	45	233
Hardware Stores	7	539

As the table shows, since the dataset has only 1 observation for "Drinking Places (alcoholic Beverages)," I may not consider "Drinking Places (alcoholic Beverages)" as an observation having impact on our result. Also, as the assumptions above, our future target customers are children, so the observation is not necessary in this data analysis.

I manipulated the dataset, Table 1, to create new variables which is the following;

1. The proportion of Non-Customers by Age Range
2. The Proportion of Customers by Age Range

With the two variables above, I created two new variables as following;

1. The proportional number of Non-Customers by Age Range and by Chain
2. The proportional number of Customers by Age Range and by Chain

Finally, I created new variable, which is the proportional number of Customers by Age Range and By Chain divided by the total number of Customers by Age Range and by Chain. This new variable will imply the probability that the Customers buy drink by Age Range and By Chain. This is simply the sales rate. The summary of the sales rate by Age Range will be the following;

Summary for Sales Rate by Age Range	
Age Range	Mean
Age 0-18	22.9%
Age 19-25	70.8%
Age 26-34	67.9%
Age 35-54	44.1%
Age 55-64	37.2%
Age 65+	31.5%

The lowest average sales rate is age 0-18, which is 0.229, 22.9%. The highest average sales rate is age 19-25, which is 0.708, 70.8%.

The following table is the summary for sales rate by Chain Category.

Summary for Sales Rate by Chain Category	
Age Range	Mean
Department Stores	41.8%
Drinking Places (alcoholic Beverages)	45.0%
Drug Stores and Proprietary Stores	51.7%
Eating Places	45.7%
Grocery Stores	46.0%
Hardware Stores	43.0%

The average sales rate in “Drug Stores and Proprietary Stores” is the highest, which is 0.517, 51.7%.

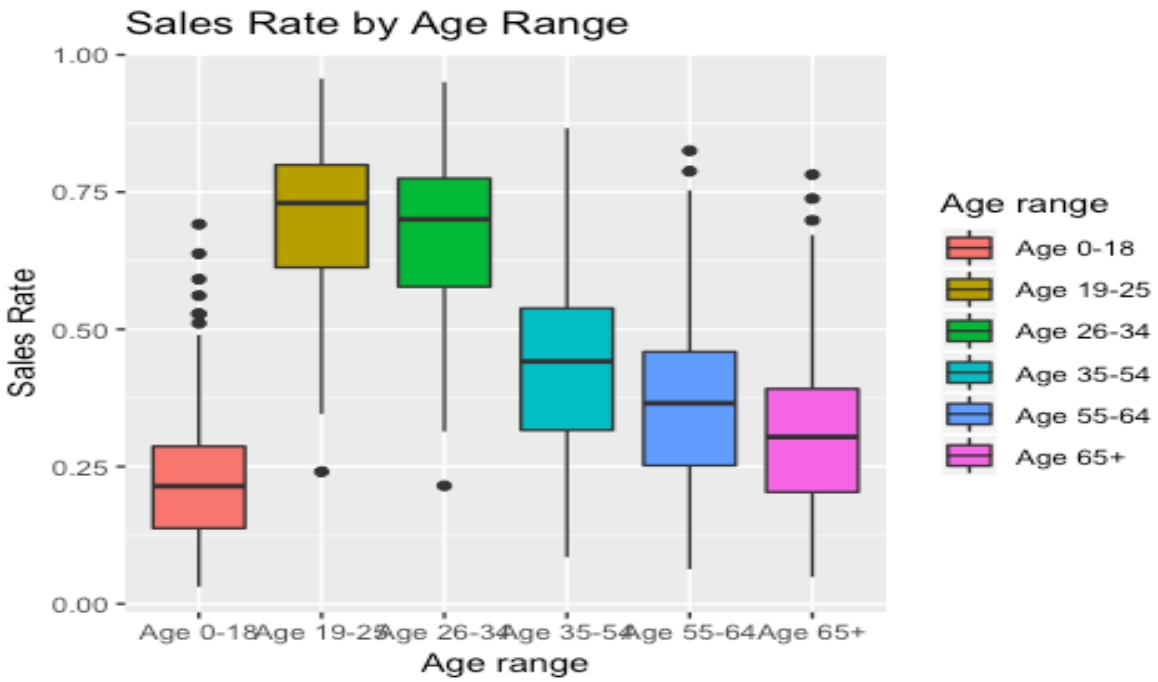
The Chain Category that has the lowest sales rate is “Department Stores.”

1. By Chain Category in Age Range 0-18

The first insights will be to find the future target chain category in age group 0-18. I will create and show several boxplots in order to easily understand the future target customer and chain category.

*Note that the line in the boxplot is not the mean of the variable, the line means the median value of the variable.

The following boxplot is for sales rate by Age Range.

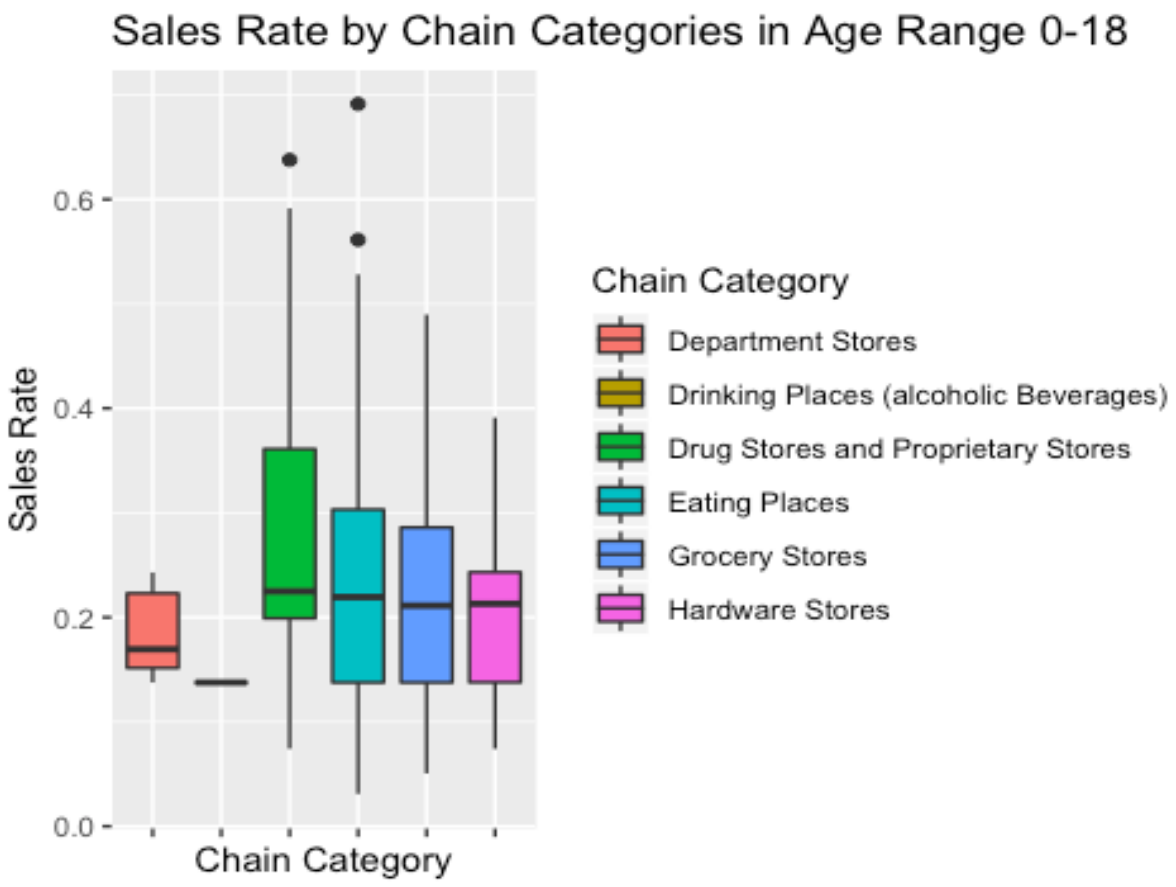


As we can see, the overall sales rate in age 0-18 is the lowest among the all Age Range. The overall sales rate in age 19-25 is the highest sales rate, not likely the Table 2 shows. Also, it will be noticed that the sales rate in Age Range 19-25 and 26-34 have the overall sales rate greater than 50%.

The following boxplot is for the sales rate by Age Range and by Chain Category.

As mentioned above, the lowest overall sales rate in Age Range is age 0-18, and the lowest overall sales rate in Chain Categories is “Department Stores.” The highest overall sales rate in Age Range is age 19-25, and the highest overall sales rate in Chain Categories is “Drugs and Proprietary Stores.” Despite the fact that this result would be because children or the customers in Age Range 0-18 don’t earn money in general, It will be better to focus on increase the sales rate for the customers in Age Range 0-18.

The following boxplot is for the sales rate by Chain Categories in Age Range 0-18, which is our future target customers;

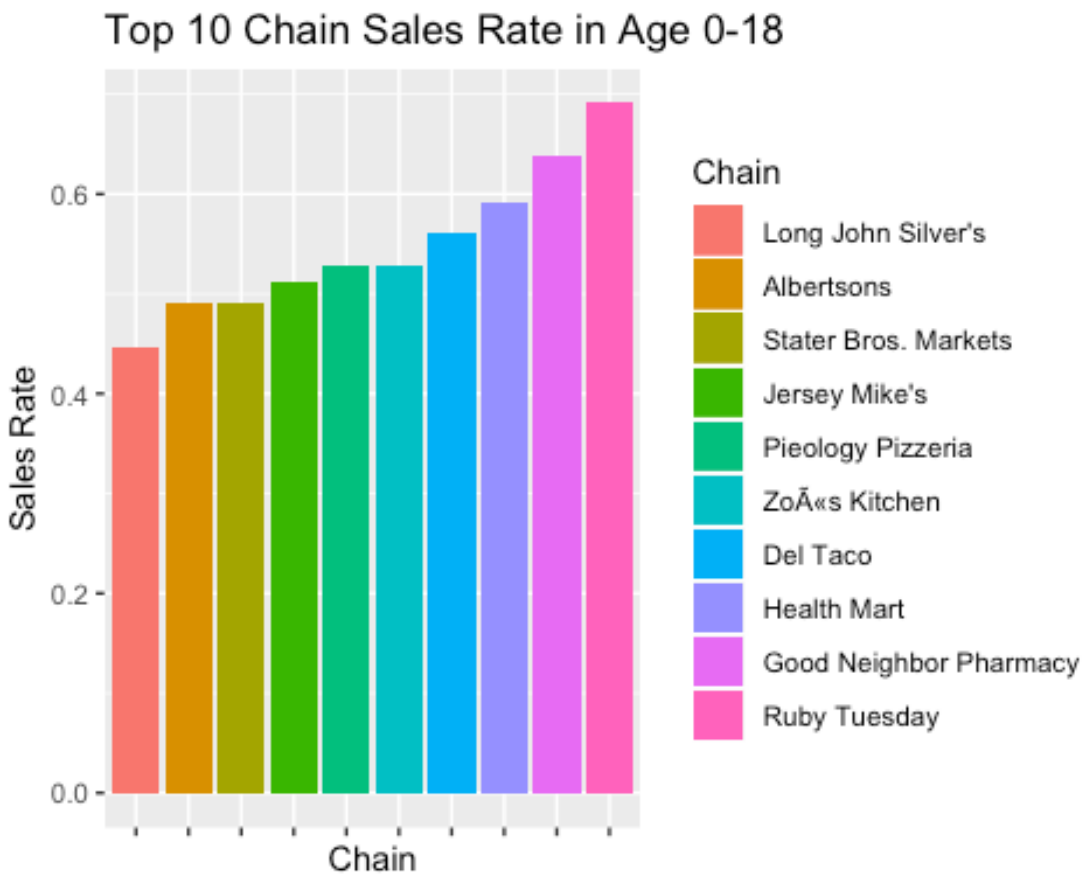


For the target customer is Age Range 0-18, the Chain Category that has the highest sales rate is “Drug Stores and Proprietary Stores.” “Department Stores” has the lowest sales rate except for “Drinking Places (alcoholic Beverages)” as the assumption above. Therefore, the future target Chain Category to increase sales rate is to focus on the “Drug Stores and Proprietary Stores.”

2. Top 10 and Worst 10 Chain for Sales Rate

I investigated top 10 and worst 10 for the probability of customers, which is sales rate, in Age Range 0-18. It will provide the information of the Chain that have top 10 sales rate and worst 10 sales rate regardless of the number of sales or customers.

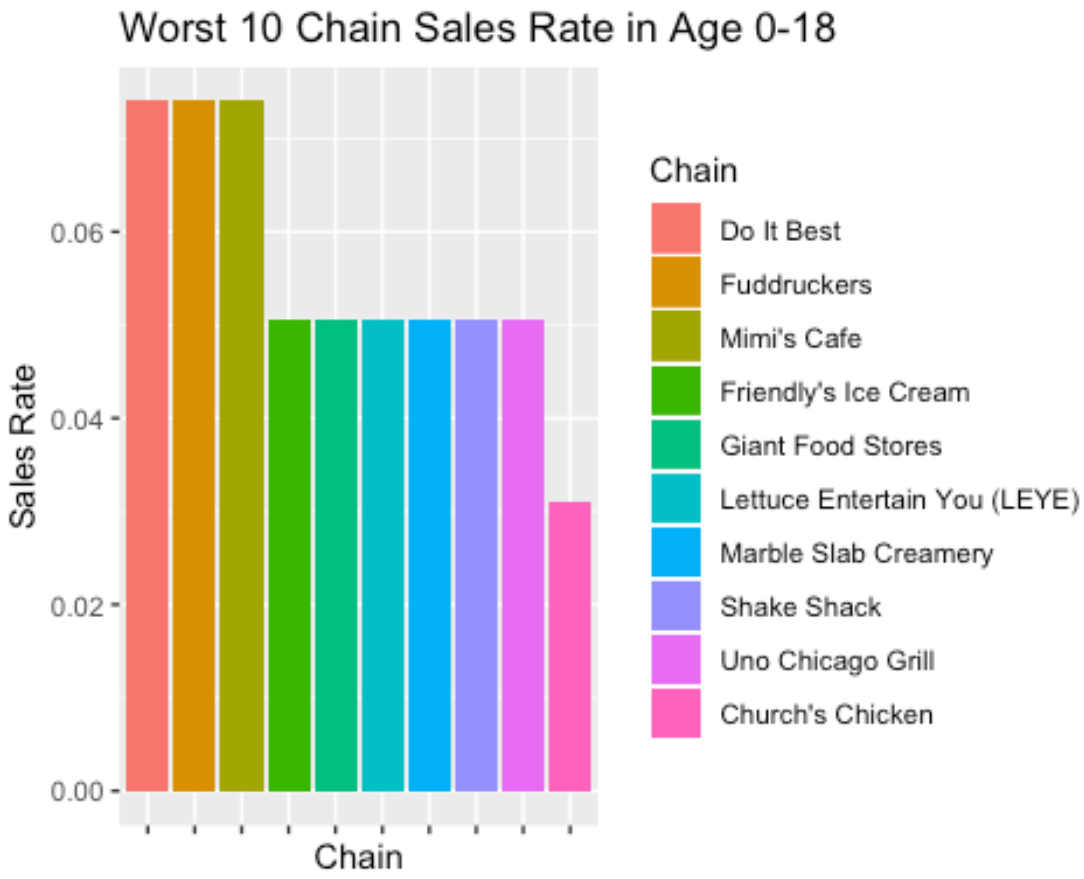
The following boxplot for top 10 sales rate in Age Range 0-18



*Note that this plot doesn't contains the information related to the number of customers.

As the boxplot shows, we can notice that "Ruby Tuesday" and "Good Neighbor Pharmacy" have the sales rate greater than 60% sales rate.

The following boxplot is for the worst 10 sales rate in Age Range 0-18



*Note that this plot doesn't contains the information related to the number of customers.

The worst 10 chains have several same sales rate, since they have the same number of customers and non-customers.

The following tables are the summary of the top 10 Chain Sales Rate and the worst 10 Chain Sales Rate. In the tables, the “Total # of Customers” is the total number of the customers in US general, which is the sum of the number of customers and non-customers.

Top 10 Chain Sales Rate		
Chain	Sales Rate	Total # of Customers
Ruby Tuesday	69.13%	82
Good Neighbor Pharmacy	63.77%	70
Health Mart	59.13%	124
Del Taco	56.12%	58
Pieology Pizzeria	52.82%	54
Zoas Kitchen	52.82%	54
Jersey Mike's	51.09%	104
Albertsons	48.98%	50
Stater Bros. Markets	48.98%	50
Long John Silver's	44.53%	91

* Note that "Total # of Customers" is the total number of customers in US General, which is sum of the number of customers and non-customers.

Worst 10 Chain Sales Rate		
Chain	Sales Rate	Total # of Customers
Church's Chicken	3.10%	131
Friendly's Ice Cream	5.06%	80
Giant Food Stores	5.06%	80
Lettuce Entertain You (LEYE)	5.06%	80
Marble Slab Creamery	5.06%	80
Shake Shack	5.06%	80
Uno Chicago Grill	5.06%	80
Do It Best	7.40%	110
Fuddruckers	7.40%	110
Mimi's Café	7.40%	110

It can be shown that the total number of customers are relatively low in the tables. It is implying that the number of customers who buy drink in Ruby Tuesday in Age 0-18 is, 82 * 69.13% = 56.6866, which is about 57 number of customers buying drink. The number

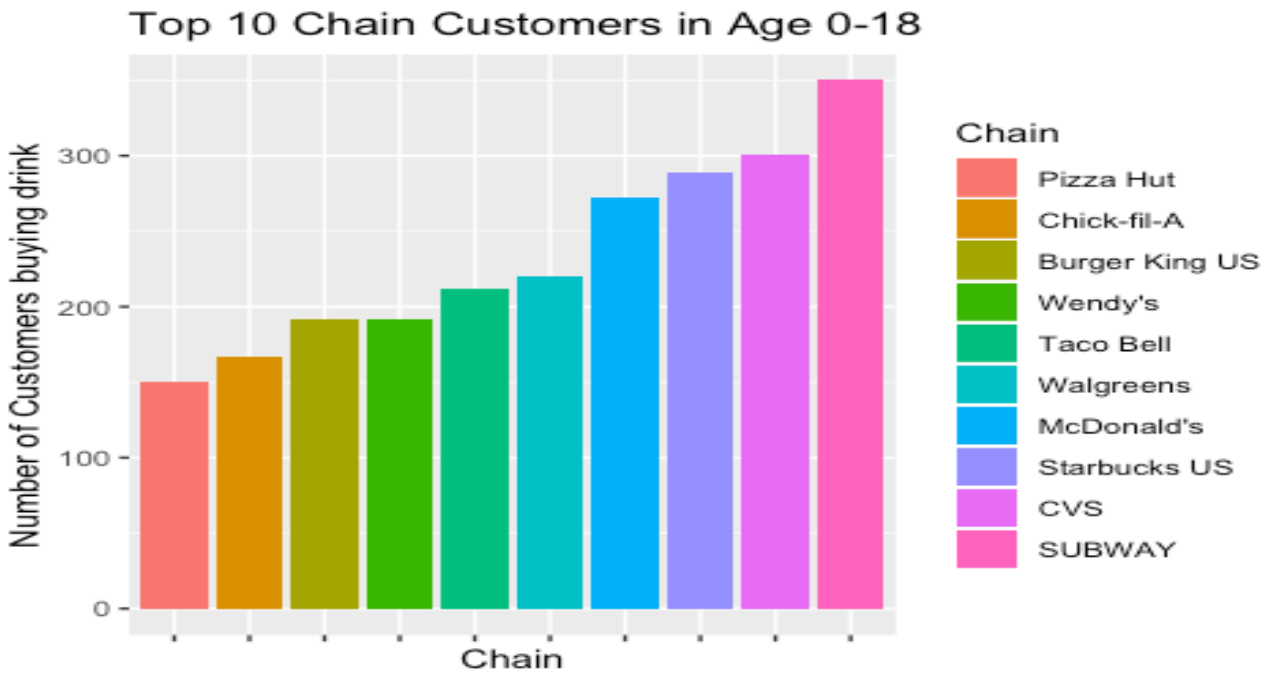
of customers who buy drink in Church's Chicken in Age 0-18 will be, 131 * 3.10% = 4.061, which is about 4 number of customers buying drink. Even though the number of customers is relatively low as the tables show, it is better to focus on top 10 chain for sales rate to increase overall sales rate.

However, we should not ignore the number of customers buying drink instead of the sales rate. Investigating the top 10 and worst 10 Chain for the number of customers will provide another insight to increase the sales volume.

3. Top 10 and worst 10 chain for the number of customers

As we can notice the plots and tables in top 10 and worst 10 chain for sales rate, the total number of customers are relatively low. Therefore, it would be necessary to investigate the top 10 and worst 10 for the number of customers.

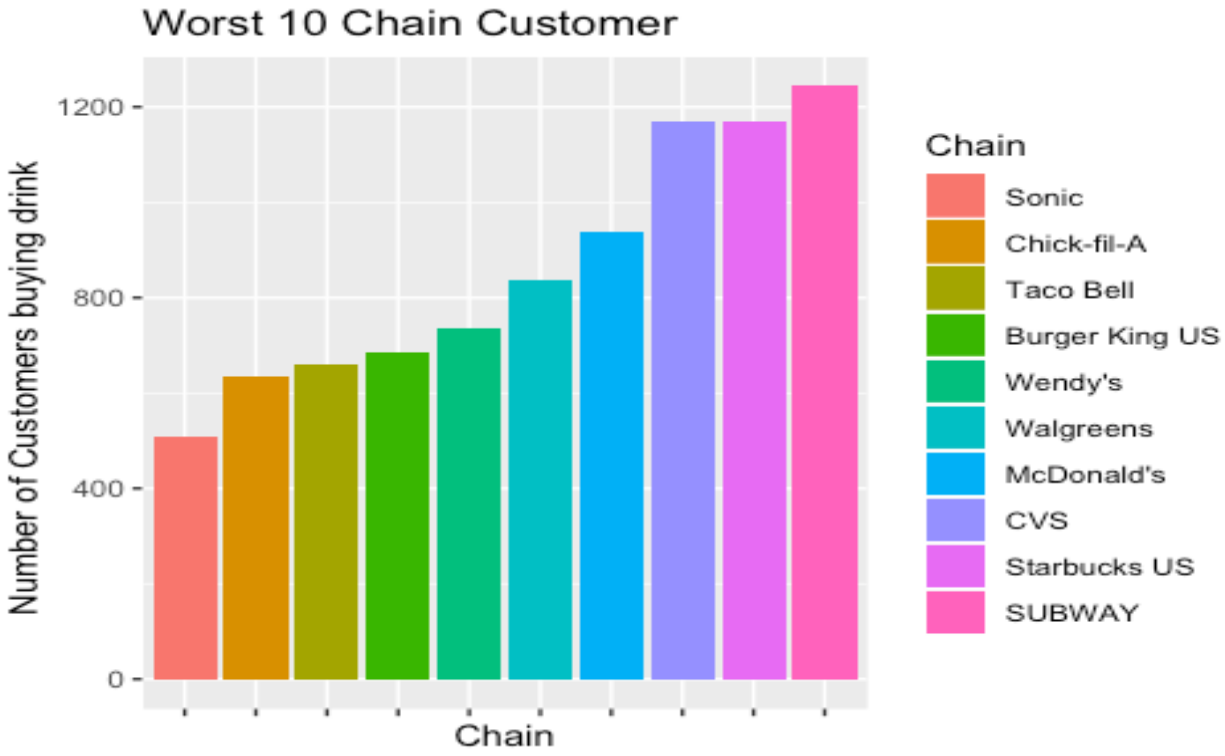
The following boxplot is for the top 10 chain for the number of customers in age 0-18



*Note that this plot doesn't contain information about sales rate.

As the plots shows, the top 10 chain for the number of customers are different with the top 10 chain for the sales rate. "Subway" and "CVS" have greater than 300 number of customers who buy drink.

The following boxplot is for the worst 10 chain for the number of customers in age 0-18



We can notice that some chain in the worst 10 for the number of customers are in the top 10 for the number of customers.

Top 10 Chain # of Customers		
Chain	# of customers	Sales Rate
SubWay	350	21.96%
CVS	301	20.50%
Starbucks US	289	19.84%
McDonald's	273	22.50%
Walgreens	220	20.79%
Taco Bell	212	24.28%
Burger King US	191	21.83%
Wendy's	191	20.63%
Chick-fill-A	167	20.82%
Pizza Hut	151	33.08%

*Note that Chain order is based on Rank.

Worst 10 Chain # of Customers		
Chain	# of customers	Sales Rate
SubWay	1244	21.96%
CVS	1168	20.50%
Starbucks US	1168	19.84%
McDonald's	939	22.50%
Walgreens	837	20.79%
Wendy's	736	20.63%
Burger King US	685	21.83%
Taco Bell	660	24.28%
Chick-fill-A	635	20.82%
Sonic	508	19.91%

*Note that Chain order is based on Rank.

We can notice that the chains of top 10 and worst 10 for the number of customers are different with the chains of top 10 and worst 10 for the sales rate. We also can notice that some chains are in the top 10 and worst 10 for the number of customers and for the sales rate, such as “Subway,” “CVS,” “Starbucks US,” “McDonald’s,” “Walgreens,” “Taco Bell,” “Burger King US,” or “Chick-fill-A.” This result is from the fact that those chains have the most number of customers. Therefore, regardless of sales rate, since they have the most number of customers, they are accounting for top 10 and worst 10. This insight will be helpful to increase the sales volume.

Conclusion

As the description states, the client company, Brawndo, is interested in some business insights for their future target customers. I developed some assumptions before starting data analysis in order to avoid confusion of understanding the dataset, such as understanding of variables or business assumption, which is that I focused on the customers in Age Range 0-18 as our future target customer, since our client company is "children's electrolyte drink company."

To make some meaningful insights that is not basic as just table shows, I combined the two tables and created new variable, which is sales rate by Age Range. With the sales rate variable, I made 3 business insights by implementing exploratory data analysis, so called EDA, by creating several plots and summary tables as the following;

1. By chain category in age range 0-18.

Despite the fact that the future target customer is the customer in age 0-18, the overall sales rate in age 0-18 is the lowest among the all age range. "Drug Stores and Proprietary Stores" has the highest sales rate for the customers in age 0-18, and "Department Stores" has the lowest overall sales rate. It would be better to expand the age range for the target customer and to focus on increasing sales in the chain categories that have high sales rate.

2. Top 10 and Worst 10 for Sales Rate

Top 10 chains for the sales rate have mostly greater than 50% of sales rate, but they have relatively low total number of customers including the customers who don't buy drink. Worst 10 chains have less than 10 % of sales rate, and they also have relatively low total

number of customers. This information implies that it will be better to focus on top 10 chains to increase sales rate, but it doesn't mean that focusing on top 10 chains increase sales volume.

3. Top 10 and Worst 10 for the Number of Customers

Top 10 chains for the number of customers have mostly greater than 200 number of customers, but they have fairly low number of customers and sales rate. Compared to the top 10, worst 10 chains for the number of customers have quite bigger number of non-customers and low sales rate. Also, most of chains in top 10 are in worst 10 as well. This result can be drawn from the fact that those chains have a large number of customers compared to others and the customers in age 0-18 have the lowest sales rate. As a result, it will be better to focus on the top 10 chains for the number of customers to increase sales rate.