Food Retail Analysis and Strategy - Mock Case

Presented by Patcharanat P.

Situation: Food Retail

Objective: Find / Extract actionable insights for improving retail business

Dataset:

- "dh_causal_lookup.csv"
- "dh_product_lookup.csv"
- "dh_store_lookup.csv"
- "dh_transactions.csv"

Tools:

- PowerBI
- Python Notebook

What proudly represented . . .

- Selected appropriate marketing campaigns and implement them based on historical data.
- Utilized Machine Learning for retail business use cases with market basket analysis (association rule).
- Identified customers' spending behavior and preferences to point out important notices in sales.
- Summary Insights, Recommendations, KPI

Table of contents

Before Analysis

Tools Selection

Since the datasets are relational databases, PowerBI seems to be appropriate to handle the data for data modeling. However, some sort of information would be more extractable by statistical analysis, so Python was used in some parts because of its flexibility.

Data Interpretation

One of the most important aspects of data analysis is understanding what data is collected and its definition, so analysts can make it useful and more interpretable. However, the definition of the attribute is not provided, so some portion of the data will be dropped off because they can't be interpreted and is not useful for this situation.

Remaining Dataset:

- "dh_product_lookup.csv"
- "dh_transactions.csv"

Data Cleaning

After observing the dataset which *seems* to be cleaned containing no missing value, a column "product_size" contains invalid values and inconsistent units such as '########', '1 LB 16 OZ', '12OZ', '2 LB', '32 OZ', or 'P 16 OZ'. So, to make it usable I cleaned the column and also use it in the analysis.

To see more detail about cleaning, please read "test_notebook.ipynb"

Cleaning processes

Filtered only valid values and valid units, and replace the others with missing values (np.nan)

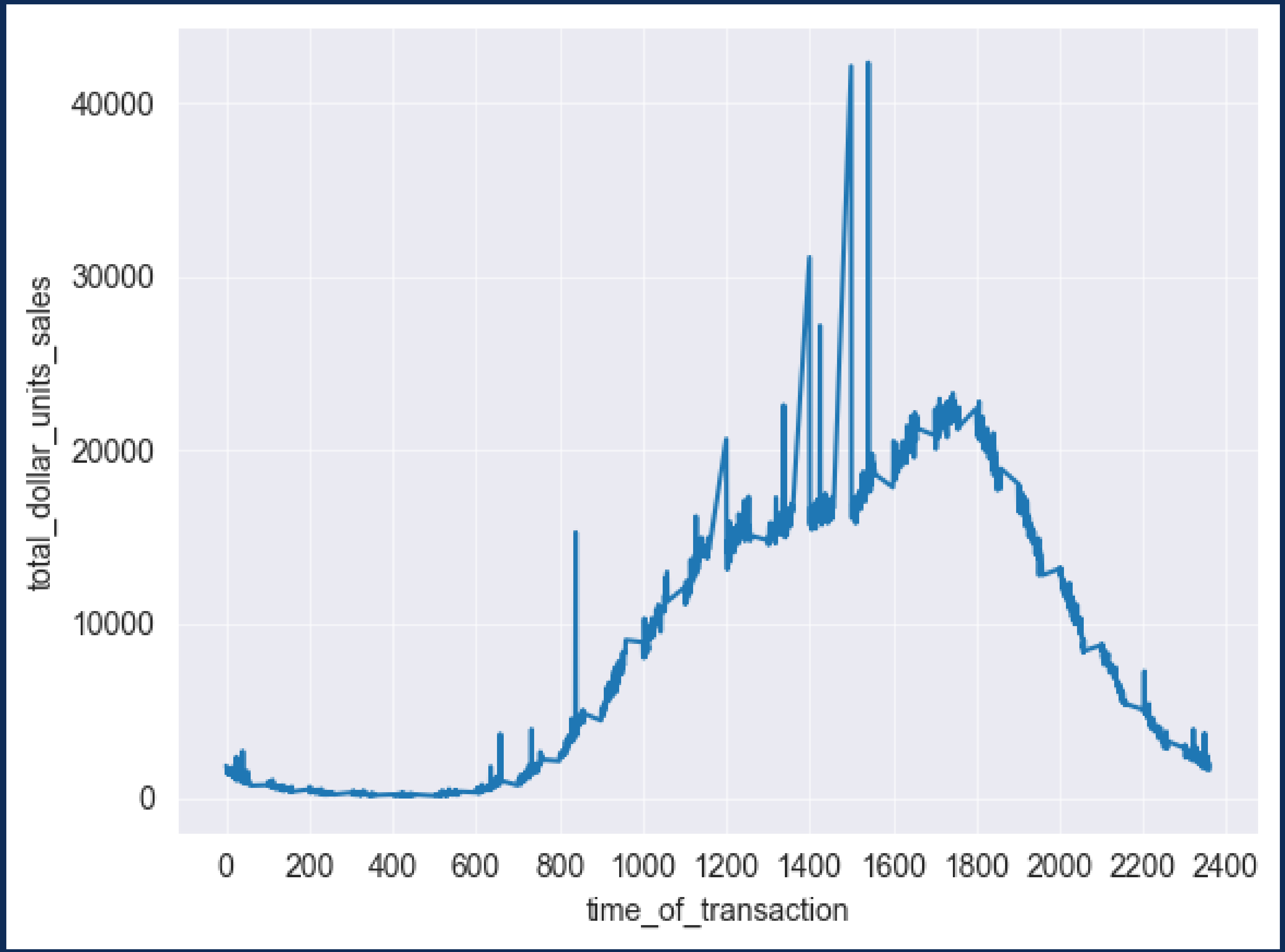
↓

Observe the column's distribution for selecting an appropriate value to fill

↓

Fill missing values with appropriate approaches

Data Analysis 1: Sales through a day



Interpretation:
time_of_transaction indicated the time in a day which allowed us to see a trend in sales.

From the graph, we can tell that people are likely to come to buy from 6 A.M. and gradually increase, having peak hour around 8 to 9 A.M., 2 P.M. to 4 P.M. Then after 6 P.M. sales gradually dropped meas people came lesser after this point.

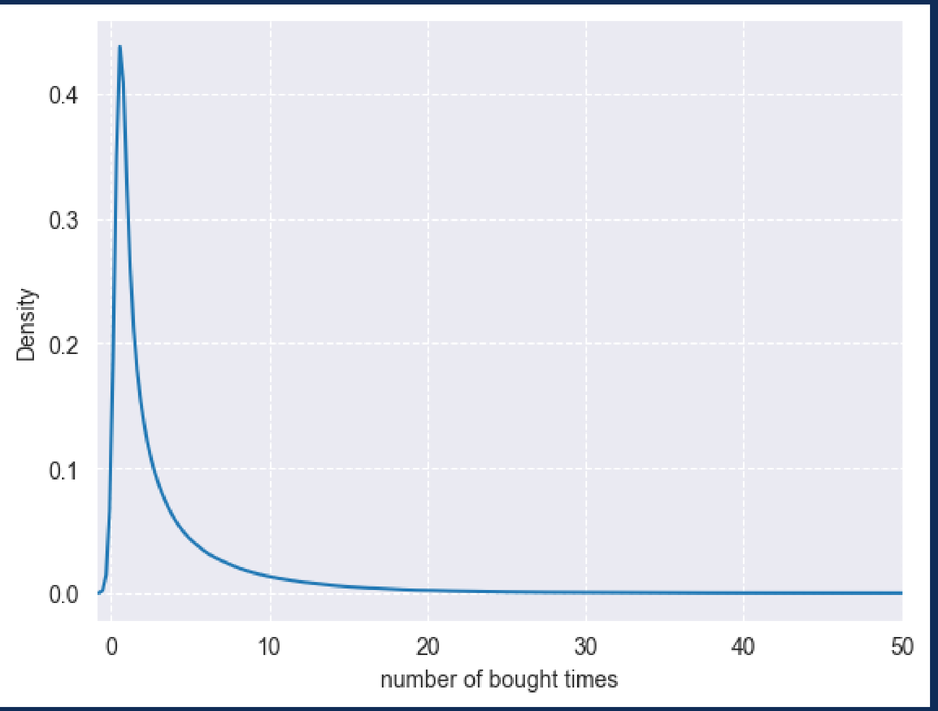
Recommendation:
I suggest applying the idea of **'food booths'** during peak hours which would get the most attention from people who are shopping.

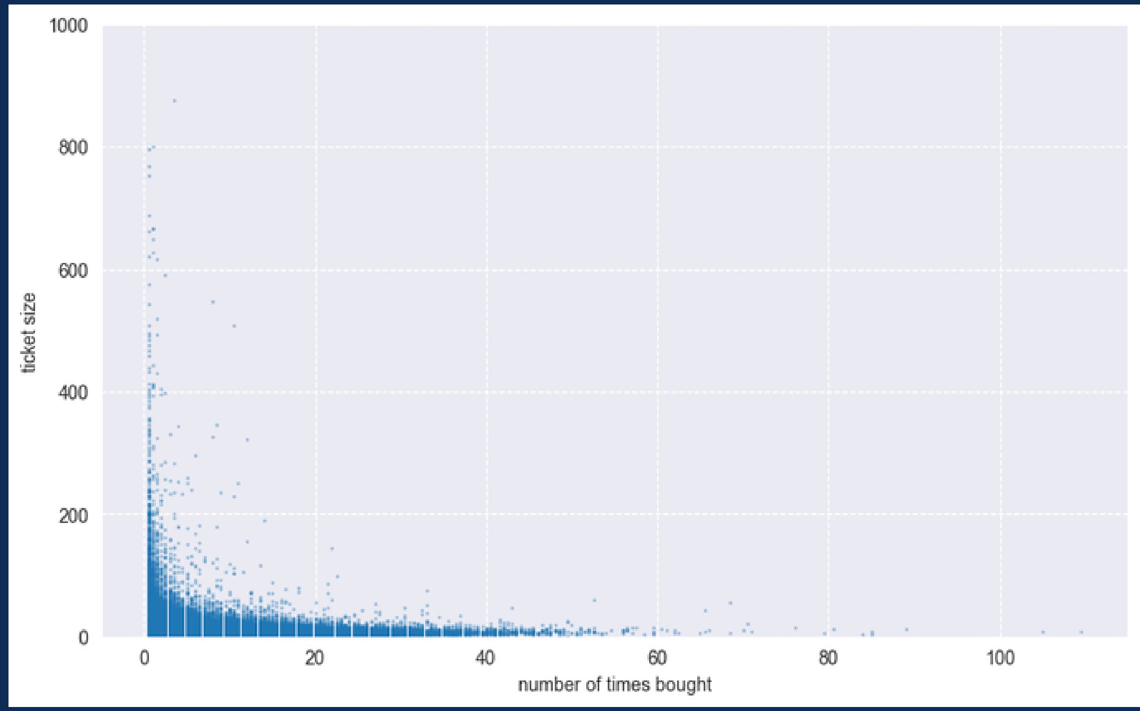
Data Analysis 1: Sales through a day

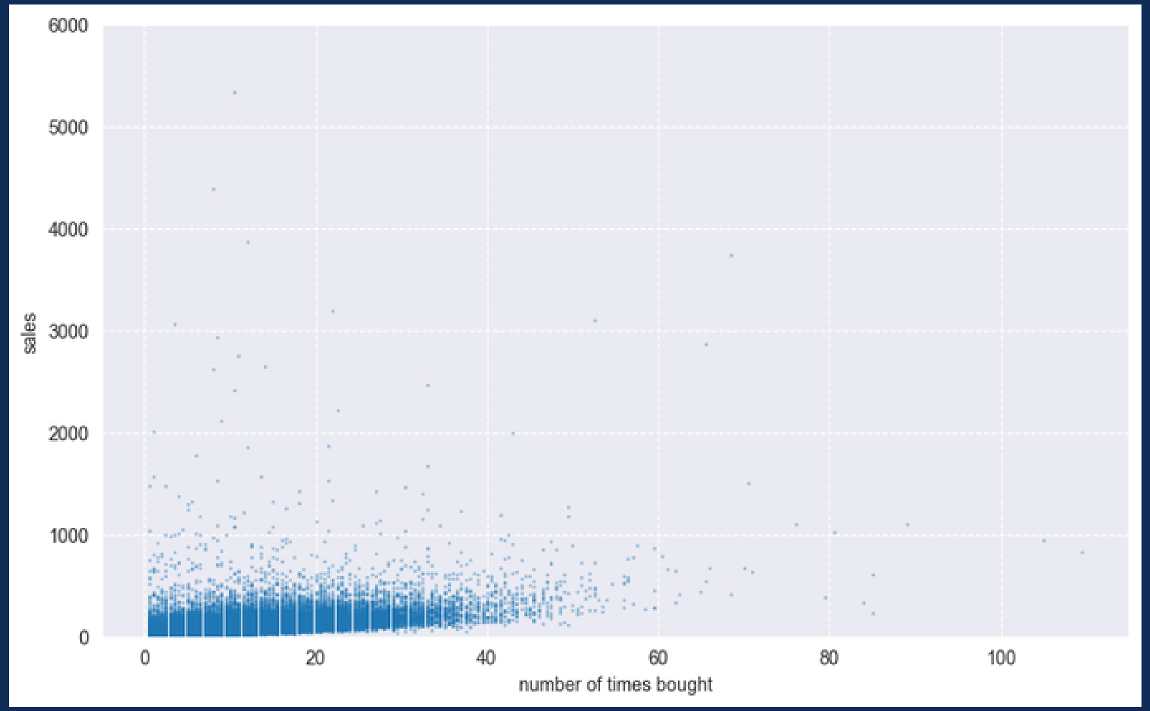
Example of food booths inside retail store

source: https://insideretail.com.au/news/woolworths-launches-labelling-program-201705

Data Analysis 2: Customers' spending behavior





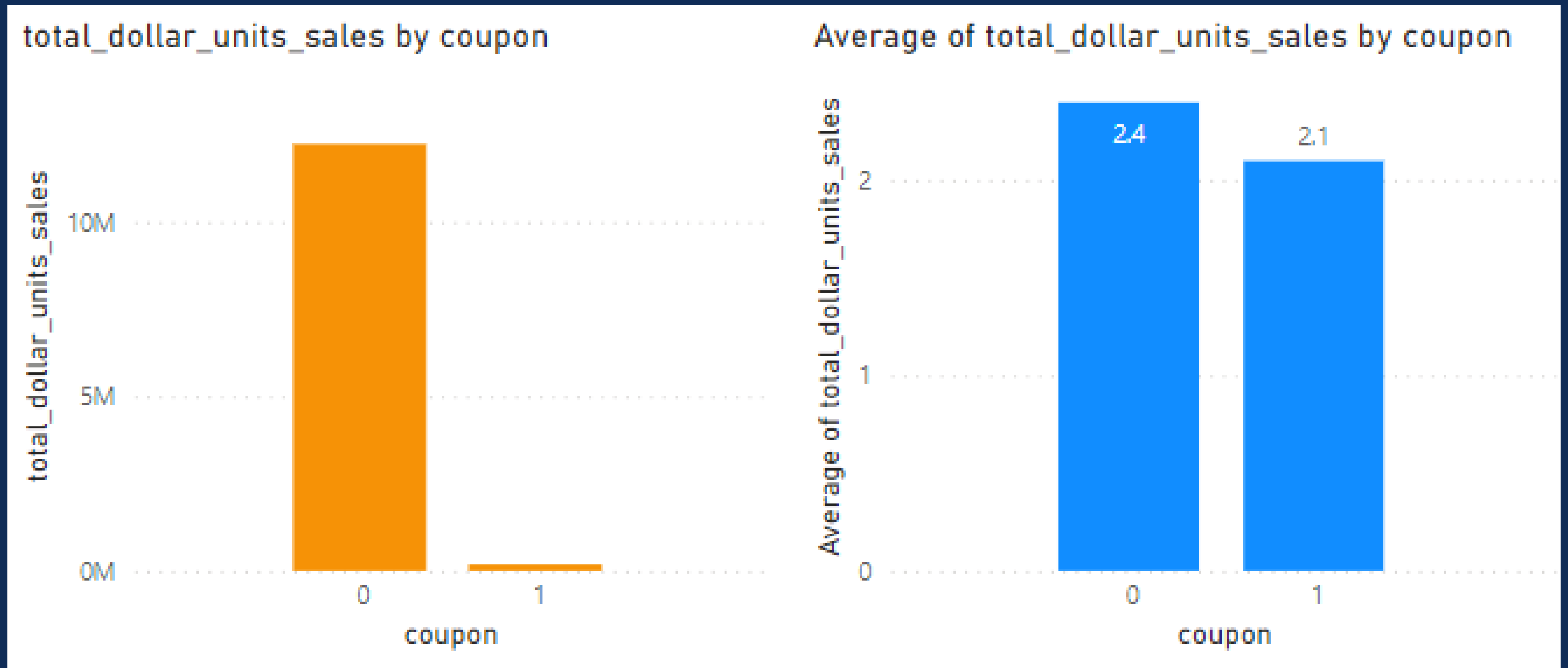


From the graph, **most people didn't come to buy often.** The average number of times people came to buy per year is 3.25 (STD 4.46). Let's say a retailer who sells groceries thing, people come to buy groceries 3 times a year is relatively low. So, **the retailer needs to increase the times people come,** In other words, to increase their loyalty.

The plot shows customers' spending behavior. We can see that **the more times they spent, the lesser ticket size they did** *(ticket size = the total amount a customer spent in one visit)*. Therefore, there are two types of customers we can classify from the graph, the first group came to buy often but spent less per time, and the second group came to buy rarely but did a big spend.

The plot shows that **the more times people came to buy the more sales increased slightly**. This implies that the retailer should encourage more times people coming to buy even if they could get a smaller ticket size referring to the previous plot

Data Analysis 2: Customers' spending behavior



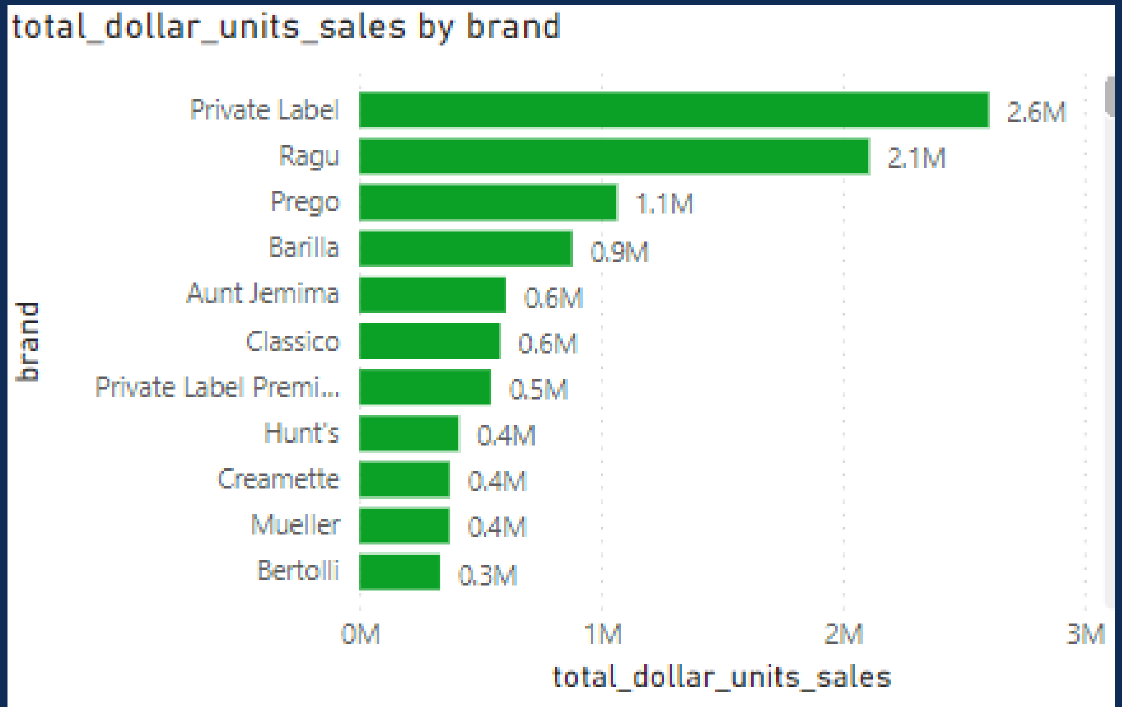
From the chart, we can see that transactions with a coupon had average sales slightly lesser than transactions without the coupon. However, total sales mostly came from transactions without coupons. This means the retailer usually didn't give the coupon to their customers, or the customers didn't usually use the coupon.

From the Previous slide and the current bar chart, I suggest that **the retailer should encourage people to buy more often even the retailer could get a smaller ticket size but in return, they could get better overall sales and also better 'Customer Lifetime Value' that associated with customers' loyalty to the retail business.**

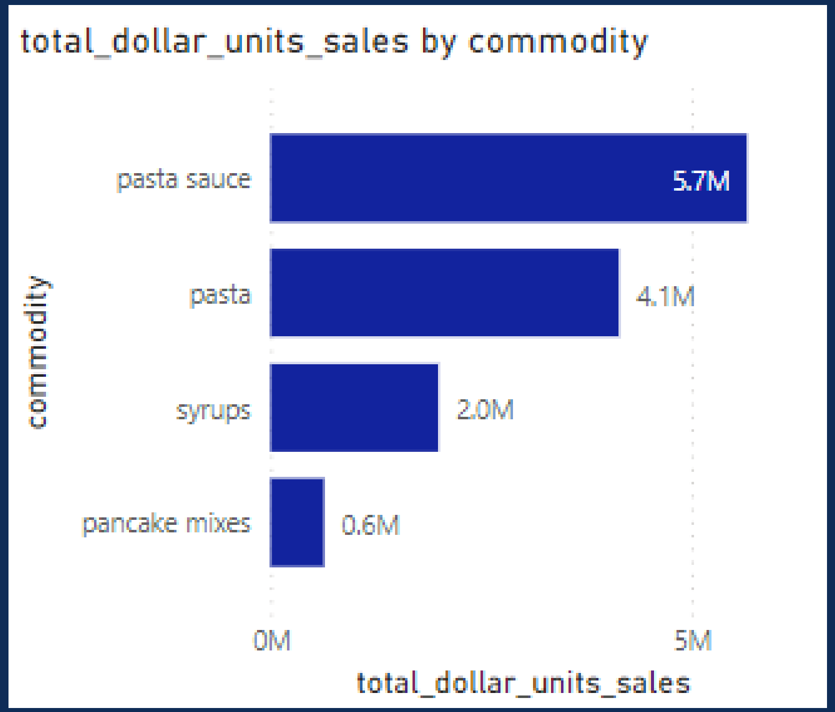
People could buy more often by giving them a **'discount next-time purchase'**. Although giving them a discount could slightly decrease the retailer's profit, the retailer would get loyalty from customers making their business become sustainable.

The bar chart shows that people didn't usually use coupons. If we assume the retailer gave more than enough coupons to customers, **the retailer should give a discount that suits customers' spending behavior**, both those who buy often with a smaller ticket size and who buy less with a bigger ticket size, **by giving certain conditions to use a discount**.

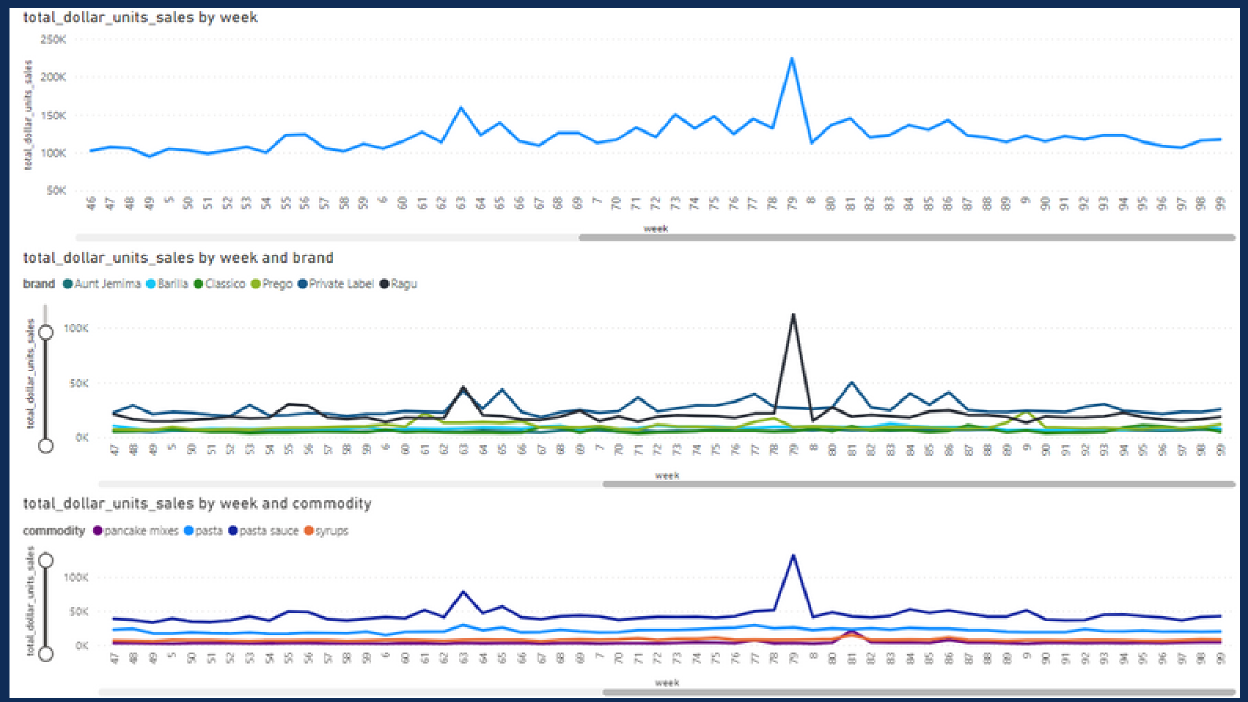
Data Analysis 3: Customer's Preference



From the chart, **<u>the most sales products</u>** came from the brands Private Label, Ragu, Prego, Barilla, and so on. **<u>This should give the retailer a sense of supplying the product properly</u>**, which one should be stocked more and which one shouldn't. Eventually, with proper stocking, the retailer could be able to efficiently save costs.



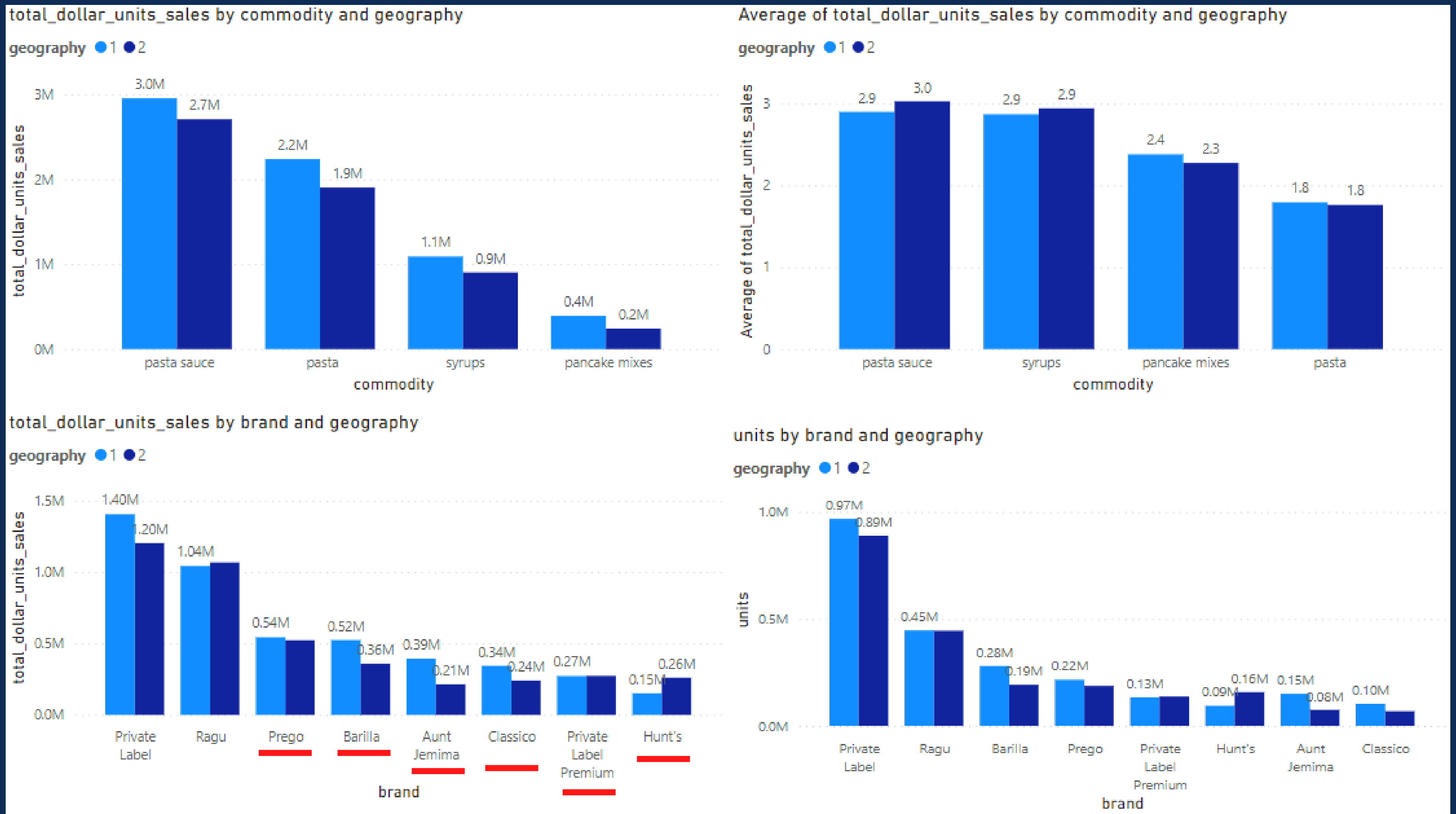
What should be noticed from the graphs is pasta sauces made the most sales to the retail store. To increase pasta sales, **the retailer should make <u>pairing promotions</u> such as <u>getting discounts when buying pasta sauce and pasta together</u>.**



From the plot, we can see that every time the brand Ragu significantly increased its sales (I guessed it was a time the products has a discount), overall sales of the retail store increased also. Unlike Private Label, the brand didn't influence customers as well as Ragu did.

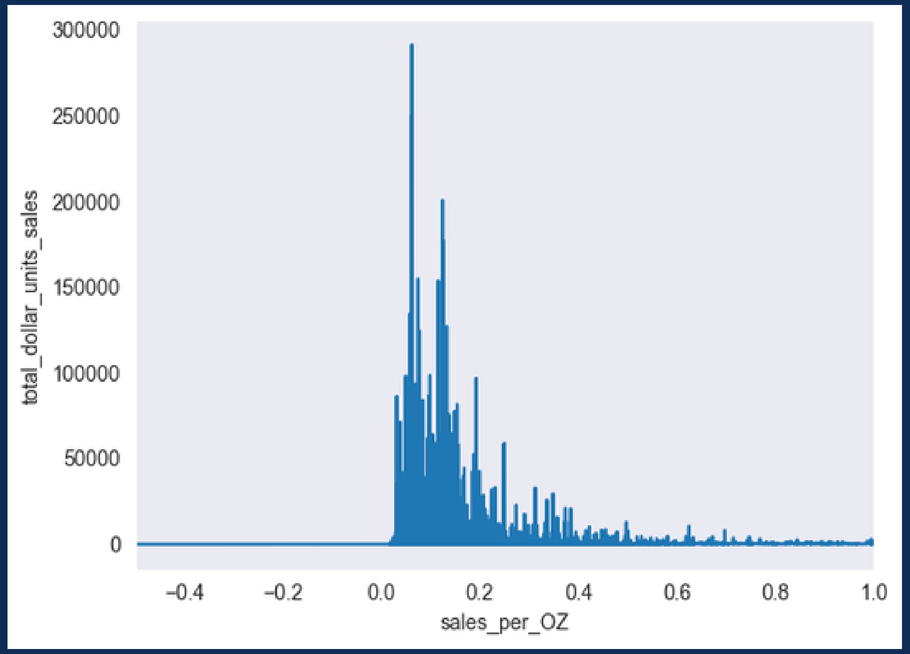
Moreover, when pasta sauce sales increased, other commodities didn't increase sales also. This could encourage pairing promotions to be made.

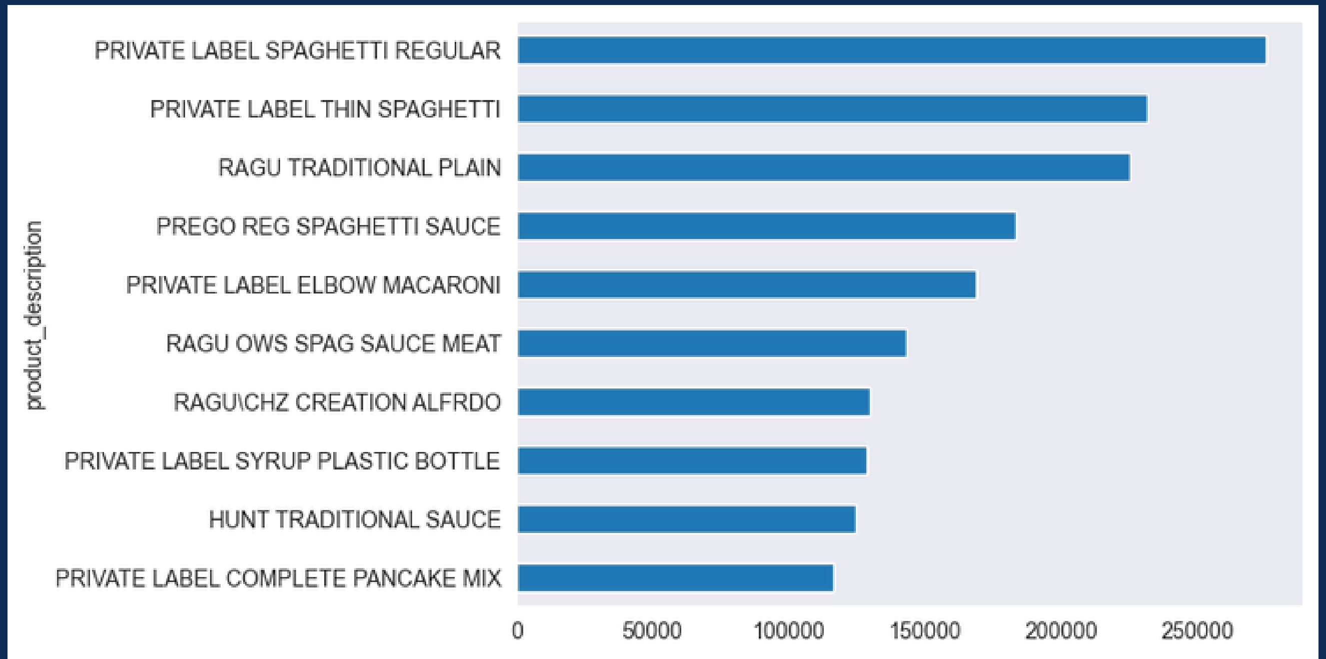
Data Analysis 3: Customer's Preference

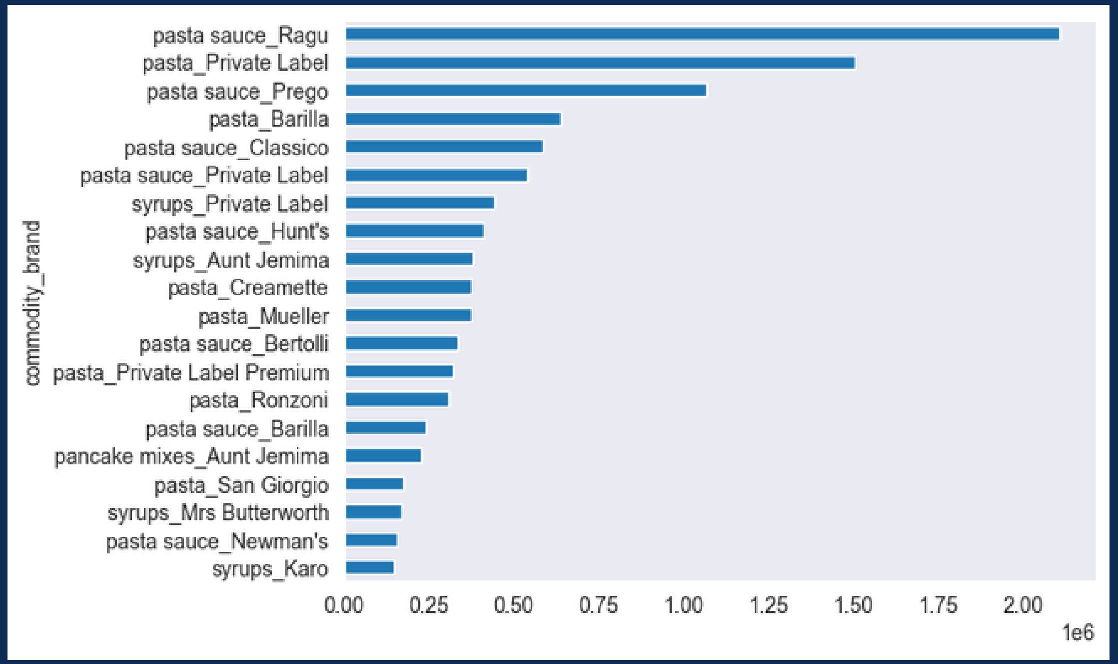


Similar to the previous slide, the graph that is highlighted could give the retailer a sense of **stocking proper brands that are demanded by different geography (and different store)**. This **could help the retailer decrease "Dead Stock"** and increase sales if they supply products and brands properly.

Data Analysis 3: Customer's Preference







The cheaper products were more likely to be bought, but there were exceptions for some products which might depend on the loyalty of the customers to the brands.

The chart shows the top 10 products that got the most sales. The retailer could use this information for **marking the products "the best seller" which could encourage the products to be bought.**

The chart shows which type of product in a brand is popular. For the Ragu brand, people usually bought their pasta sauces. For the Private Label, people usually bought their pasta, and so on.

Data Analysis : Market Basket Analysis

antecedents	consequents	antecedent support	consequent support	support	onfidence	lift	levera
(pasta sauce_Private Label)	(pasta_Private Label)	0.107075	0.423098	0.066752	0.623414	1.473450	0.0214
(pasta_Private Label)	(pasta sauce_Private Label)	0.423098	0.107075	0.066752	0.157769	1.473450	0.0214
(pasta_Private Label)	(pasta sauce_Hunt's)	0.423098	0.086188	0.046711	0.110403	1.280951	0.0102
(pasta sauce_Hunt's)	(pasta_Private Label)	0.086188	0.423098	0.046711	0.541968	1.280951	0.0102
(pasta_Mueller)	(pasta sauce_Ragu)	0.087683	0.327852	0.036467	0.415896	1.268550	0.0077
(pasta sauce_Ragu)	(pasta_Mueller)	0.327852	0.087683	0.036467	0.111230	1.268550	0.0077
(pasta sauce_Ragu)	(pasta_Creamette)	0.327852	0.090273	0.033996	0.103694	1.148667	0.0044
(pasta_Creamette)	(pasta sauce_Ragu)	0.090273	0.327852	0.033996	0.376592	1.148667	0.0044
(pasta_Private Label)	(pasta sauce_Ragu)	0.423098	0.327852	0.149335	0.352956	1.076571	0.0106
(pasta sauce_Ragu)	(pasta_Private Label)	0.327852	0.423098	0.149335	0.455495	1.076571	0.0106
(pasta_Private Label)	(pasta sauce_Prego)	0.423098	0.162533	0.070268	0.166079	1.021817	0.0015
(pasta sauce_Prego)	(pasta_Private Label)	0.162533	0.423098	0.070268	0.432329	1.021817	0.0015
(pasta_Barilla)	(pasta sauce_Ragu)	0.136442	0.327852	0.037704	0.276334	0.842865	0.0070
(pasta sauce_Ragu)	(pasta_Barilla)	0.327852	0.136442	0.037704	0.115002	0.842865	0.0070

Market Basket Analysis can give us certain pairs of products that usually be bought together making us able to initiate a campaign or promotion for that pair of products to encourage sales.

"support score", tells us how frequently it was bought or occurred in transactions records

"lift score", could be interpreted as an associated score between products. The more score it get, the more likely they were associated with each other.

Example:

- The people who bought **pasta sauces from the brand Private Label**, are likely to buy **pasta from the brand Private Label** also.
- The people who bought **pasta from the brand Private Label**, are likely to buy **pasta sauces from the brand Hunt's** also.
- The people who bought **pasta from the brand Mueller**, are likely to buy **pasta sauces from the brand Ragu** also.

and so on...

Insights Summary

1. There's a **peak time during the day** when the retailer could get more profits if they performed the right action.
2. **Customers have different spending behavior**, **mostly spend not that often and keeping ticket sizes low**. The most important is **the more often people came to spend, the more sales increased slightly regardless of ticket size.**
3. Most customers didn't use a coupon to get benefits, and **average sales caused by customers who used the coupon and who didn't are quite equal** (slightly different).
4. There are some specific brands that customers preferred, and **pasta sauces influence customers more than other types of commodities.**
5. Different geography (and stores) have different demands for brands.
6. There are the top product sales that the retailer can benefit from by emphasizing them for the customers.
7. **Market Basket Analysis can provide certain pairs of products that usually be bought together** helping the retailer to build a campaign and promotion.

Recommendations and KPIs

1. I recommend using "Food booths" especially during peak hours of the day to get the most attention from customers and also offer them pairing promotions with certain product pairs given by market basket analysis which would increase sales of other types of commodities besides pasta. Moreover, when they get a discount or benefit, they would be more loyal to the retailer which causes them to come back to buy more often which finally will increase overall sales significantly.
2. I also recommend giving the customers a proper discount that suits their spending behavior which can be done by determining conditions to use a discount to make them buy more often with a certain minimum ticket size, such as spending a minimum of 8$ to get a 1.5$ discount the next time purchase **or** spending 25$ to get 8$ discount this time purchase.
3. The different geography and stores had different demands for brands. If the retailer could stock products properly, there will be no "Dead Stock" increasing sales for each store or geography and decreasing the retailer's costs.
4. A discount or pairing promotion on popular products of pasta sauces could lead to a significant sales increase in overall sales (including other commodities).

Food Retail Analysis and Strategy - Mock Case

Presented by Patcharanat P.
Patcha-ranat@outlook.com