Analysis of All Foods Data

Team: Data Brokers
Zach Bowyer, Logan O'Brien, Arjun Sharma, Rhea Sharma

I. Introduction/Overview

This report details the issue of declining revenues for the store 'All Foods', which is located in Melbourne, Australia. Here, we outline the dataset we used for analysis, explain our data cleaning process, discuss initial exploratory data analysis (EDA) of the provided dataset, give general insights/analytics, and lastly give some recommendations for the business. Overall, while we made a few key assumptions about the business, we found there are a few interesting facts that give valuable insight into why the business is struggling in 2018 relative to 2017.

II. Assumptions made

- The Australian economy experienced growth over 2017-2018 [1]. Therefore, general economic growth/decline will not be attributed to the negative change in profits.
- When the store changed its operating hours in 2018, the difference in shelf management, store organization, store inventory, employee wages, marketing strategy, and store cleanliness was kept constant (Realistically we know these aren't the same, but we simply do not have the information to do anything about it).
- Food generally is price inelastic (demand does not shift as price shifts). Therefore, we cannot attribute changes in the quantity bought to any economic factors.
- We are assuming that revenue only relates to sales of products (Not accounting for operations and infrastructure).

III. Data Cleaning Process

- Combined: Within the main category, the categories "Beverage" and "Beverages" are combined.
- Unit_buying and total_buying had null values. These missing values were imputed using KNN Imputer. The k-Nearest Neighbors (kNN) imputer is an imputation algorithm that can be used to fill in missing values in a dataset, using the kNN algorithm to impute missing values. It works by finding the k-nearest neighbors of each instance with missing values and taking the average (for continuous variables) or the mode (for categorical variables) of those neighbors to impute the missing value.
- The subcategories were changed as follows:
    1. Sweet to Sweets
    2. Condiment to Condiments
    3. Bananas, Apples, Pears, Tomatoes, Pumpkins,and Stonefruits combined under Fruits
    4. Papad combined under Snacks

5. Cake and Sweet Bread combined under Baking
6. Onion, Potatoes, and Garlic combined under Veggies
7. Spices, spides, and spices combined into “spices”
8. Sauce and sauces combined into “sauces”
9. Coconut products and cocunut products combined into “Coconut products”
10. packaed and packaged combined into “packaged”

- Time and Date Extractions: extracted *year*, *month*, *day*, and *hour* columns from *date* column

IV. Exploratory Data Analysis

After we performed data cleaning, we spent some time doing an initial exploratory data analysis. In this analysis, we discovered the following:

- There were about 10K fewer orders in 2018 resulting in $43,585 less in sales.
- The ‘Fresh Produce’ category is by far the highest-selling category across all months and hence an important focal point in our analysis

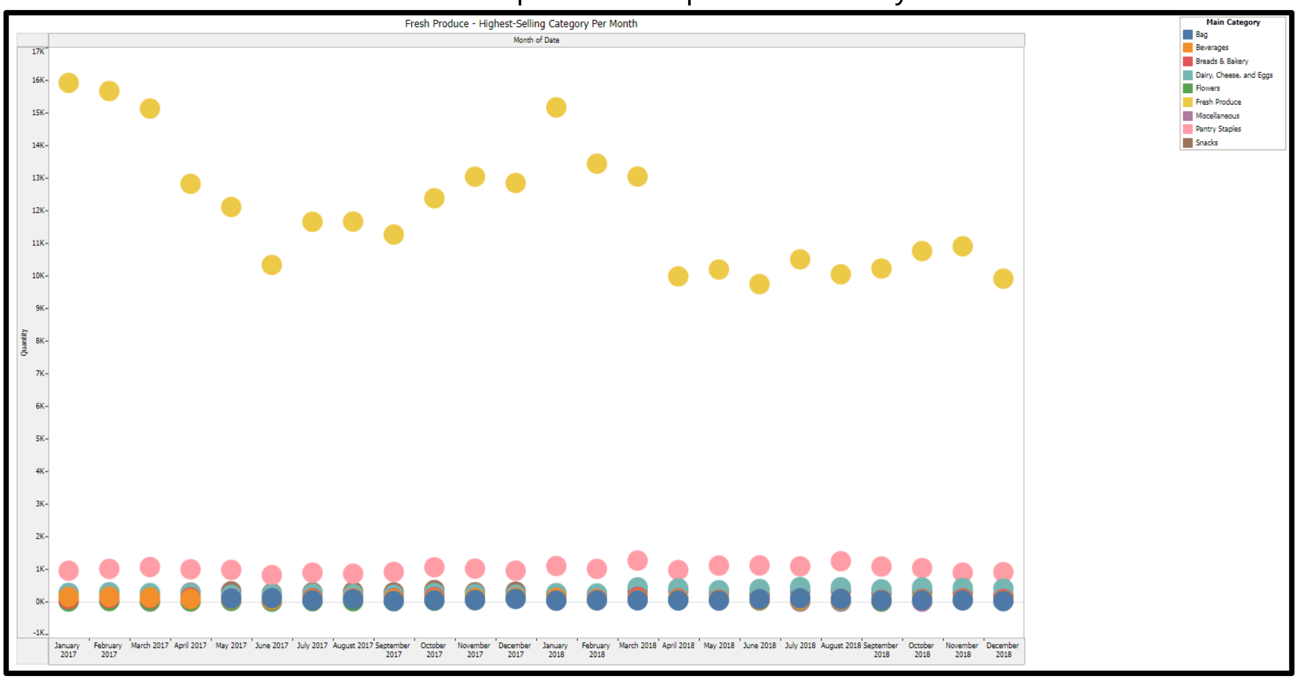

Figure: Shows how well each category sells

- Within the ‘Fresh Produce’ category there is some significant seasonal variation in a portion of the subcategories.

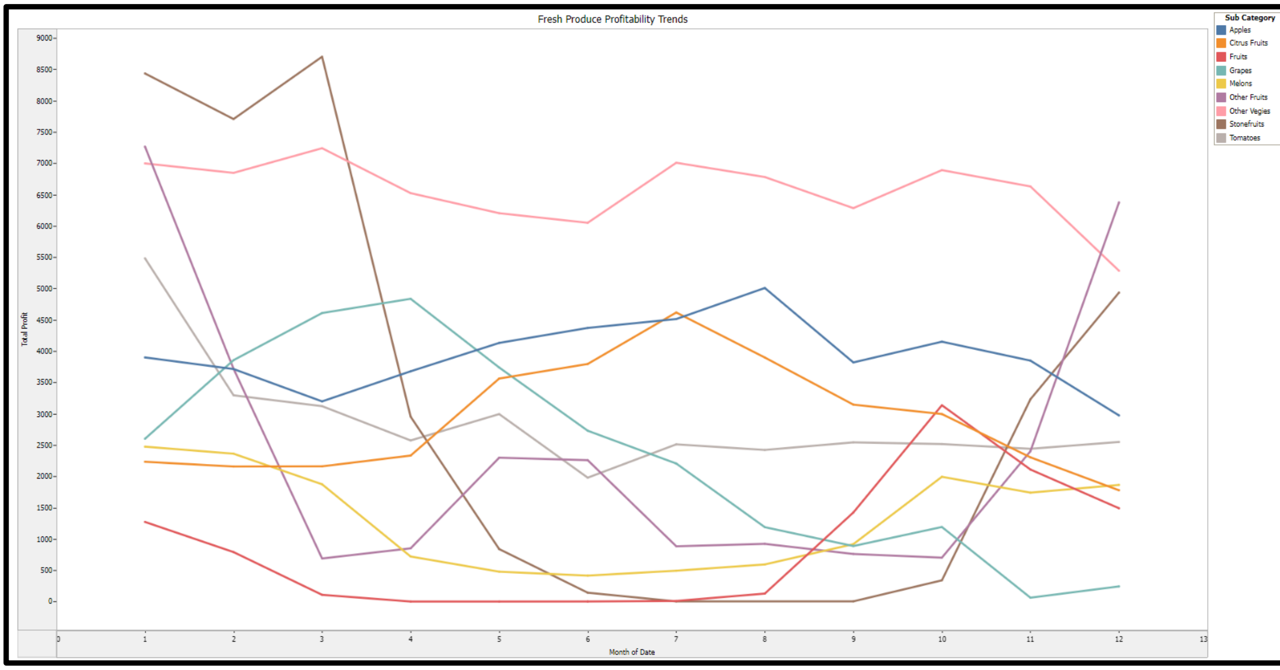

Figure: Shows variation of each category

- The profit that the store makes heavily fluctuates based on the current season. More specifically, Summer (206.53k) > Fall (179.73k) > Spring (162.47k) > Winter (156.71k). The visual below illustrates the total profit per day.

Figure: Shows the total profit of the store over the 2-year span

- Fresh produce is the overwhelming majority of sales and profits. This is due to the low spoiling time of such products. The visuals below show how most categories are relatively near-zero in total profits to fresh produce.

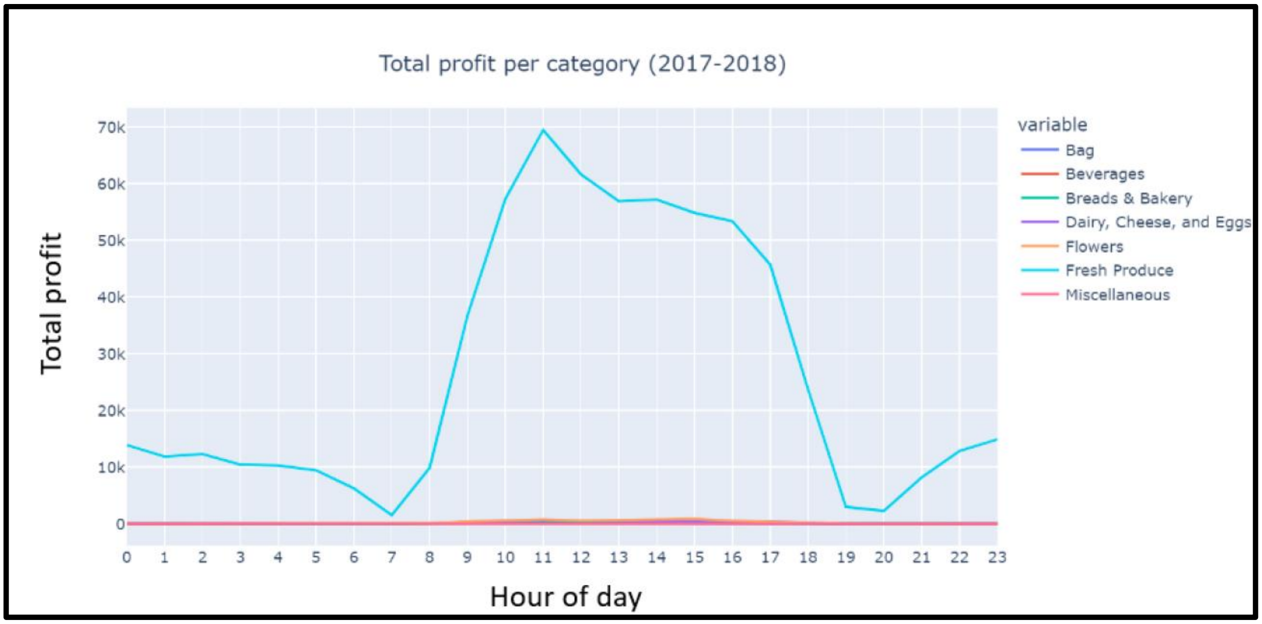

Figure: Shows the total profit of each category over the hours of the day

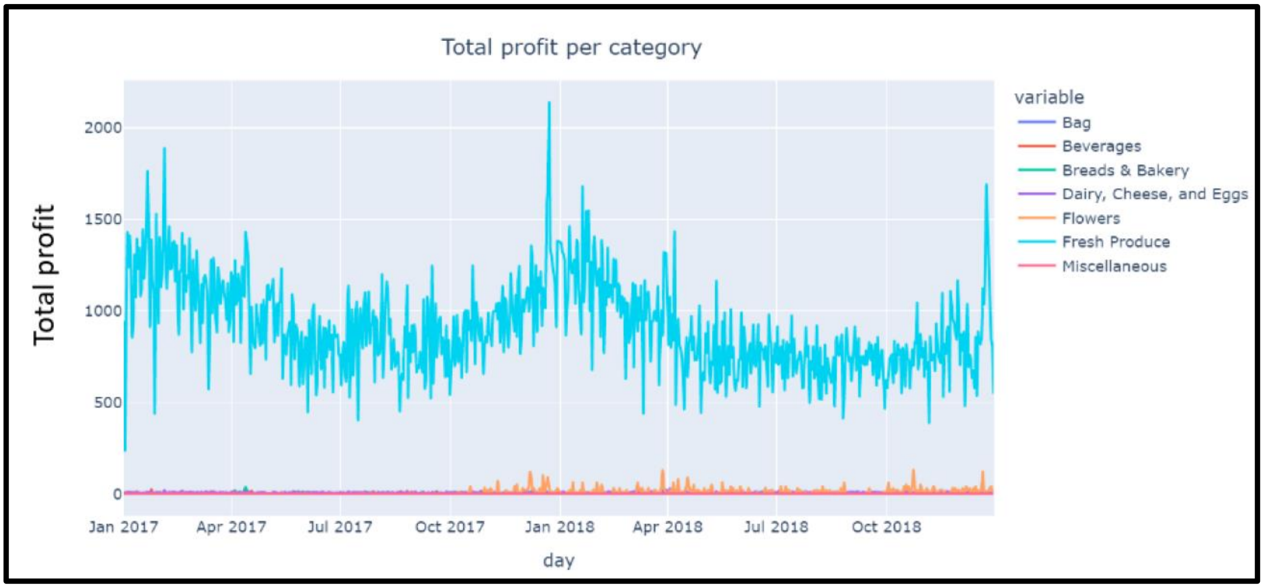

Figure: Shows the total profit of each category over the 2-year span

V. Analysis

<u>Basket Size Through the Day:</u>

The difference in the hours of operation between 2017 and 2018 warrants an individual analysis as the difference can prove to be significant. The figure below shows the Basket Size Through the Day.

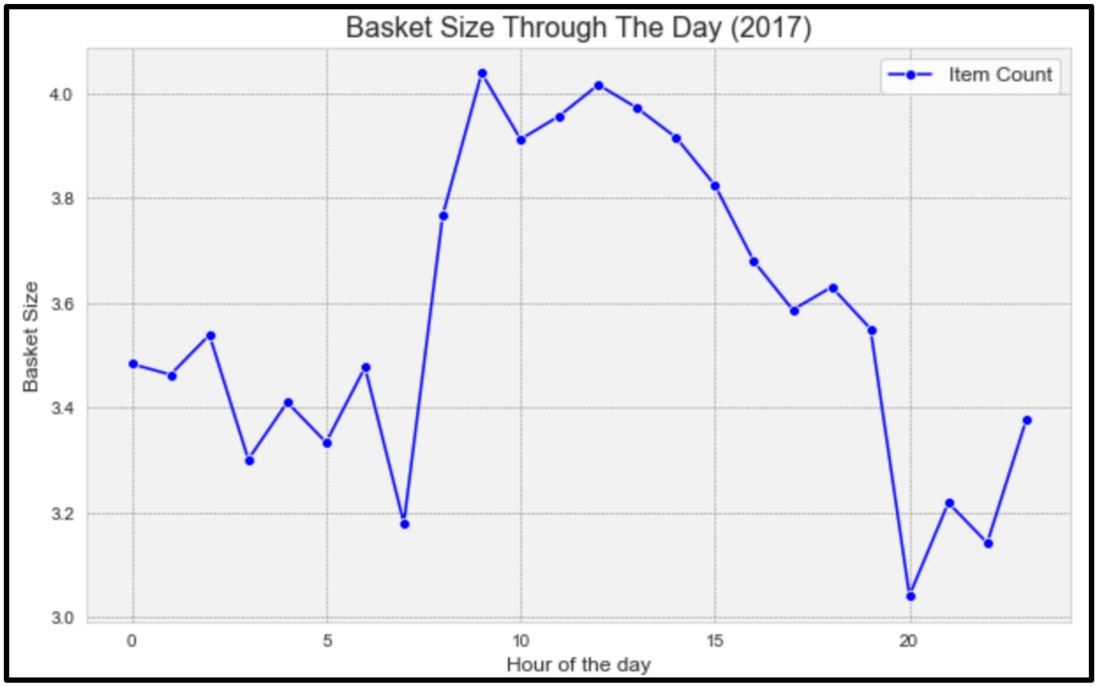

Figure: Basket Size Through The Day (2017)

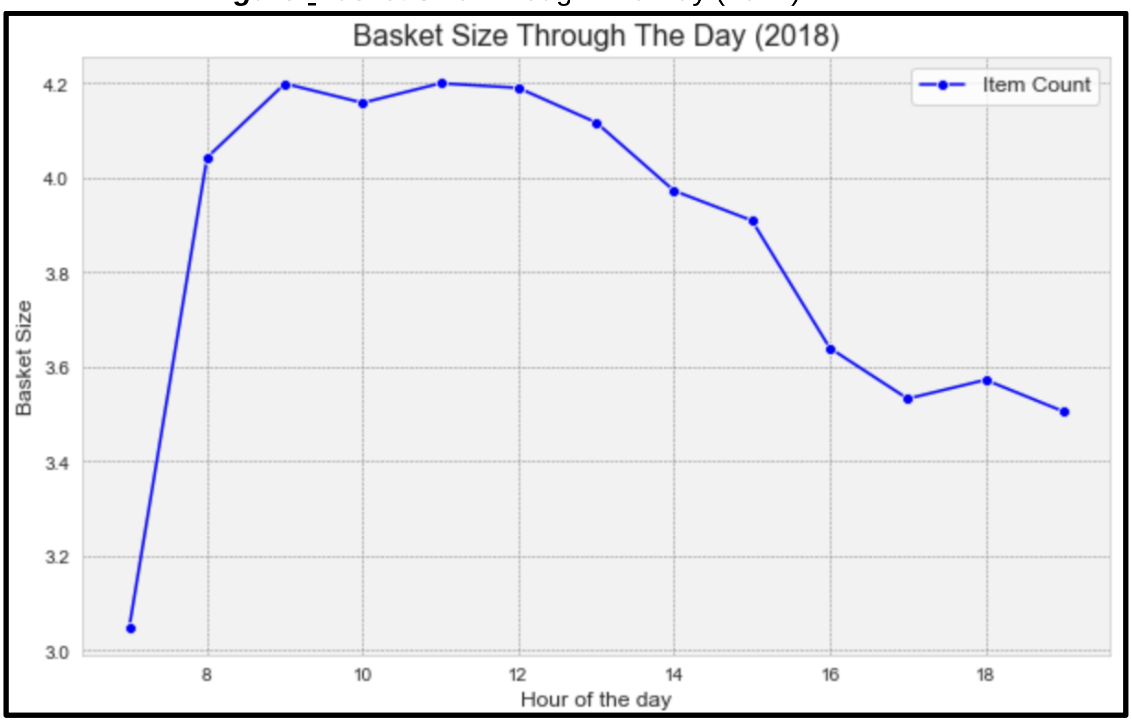

Figure: Basket Size Through The Day (2018)

It is observed that for the common hours for which the store is open, the trend of average basket size is preserved, with minor fluctuations. The average basket size increased in 2018, with the average basket in 2017 having 3.57 items, compared to 3.85 in 2018. The circumstances suggest that customers tend to buy more items to reduce their visits to the store considering the shorter time frame of store operation.

Basket Size Through the Months:

The Basket Size is visualized for the months of the year. We do not take separate plots for both years because we are interested in the trend of customer's purchases overall, and that incisive details relating to daily purchase and visit patterns would not be conducive to answering this question. We are interested in monthly trends in order to assess shopping behaviors over seasons and important events.

Figure: Basket Size Through The Months

We see that the basket size tends to increase over the months of the year. The higher basket size in the second half of the year can be attributed to increased activity on multiple fronts. The increase for the months of July to September can be explained by the annual summer vacation taken across the country by students and professionals alike. Furthermore, this time also constitutes "back to school" shopping and preparation, hence partially accounting for the surge in business. The Months of October to December are quite busy with several public holidays like Halloween, Thanksgiving, and Christmas. These holidays would prompt shoppers to prepare for the occasion, and hence would lead to specific purchases, leading to an increase in the basket size.

The initial few months of smaller baskets may be partially attributed to multiple contributors such as New Year's. Shopping for the new year was largely completed in December of the previous year. In addition, the less frequent public holidays leave little motivation for recreational purchases. It is speculated that this will also be accompanied by higher market visit frequency at the earlier times of the year as compared to the later months because customers might prefer to complete shopping for occasions along with that of their own personal requirements.

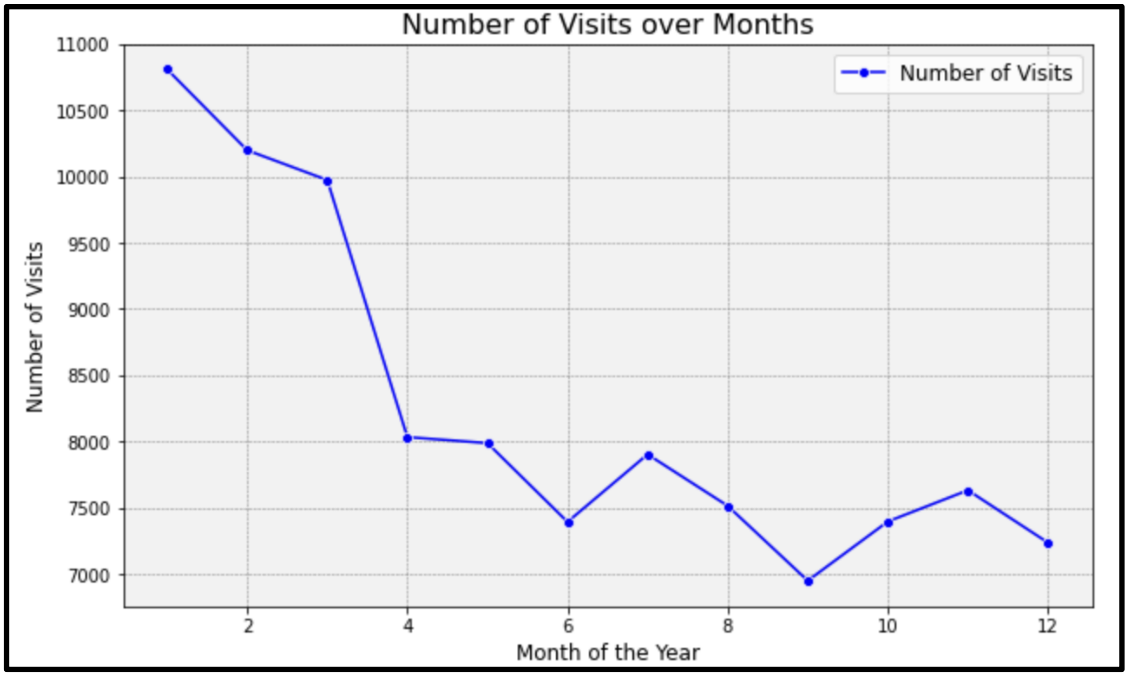

Figure: Visit Frequency Over the Months

The visual above corroborates prior speculation. The monthly trend for visit frequency is somewhat a reversal of the basket size.

<u>Trends in Transaction:</u>

There are three main methods of payment, namely “cash”, “magcard”, and “free”. However, out of the almost 100,000 orders over the 2 years we consider, only 10 orders were free. Hence, the free orders are not considered for the transaction trends.

First, we look at the distribution of the proportion of revenue incoming as a result of cash or card payments.

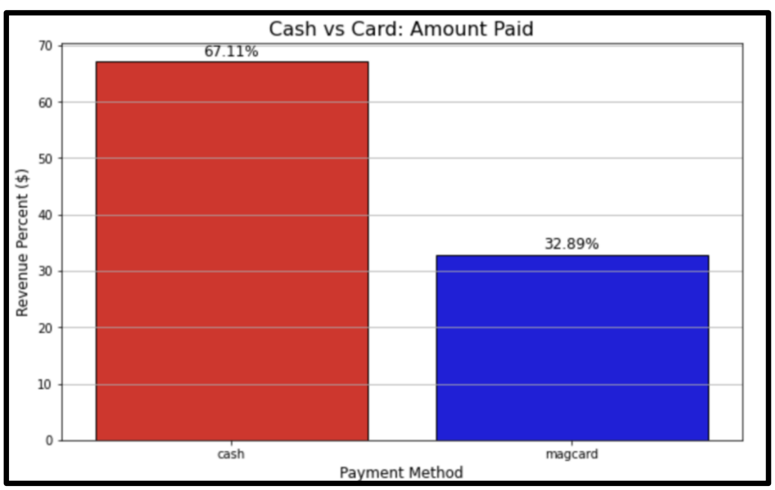

Figure: Percentage of Incoming Revenue from Cash and Magcard.

Over two-thirds of the revenue comes from cash. Card payments amount to less than a third of the incoming revenue.

Next, we look at the proportion of transactions made through each payment method.

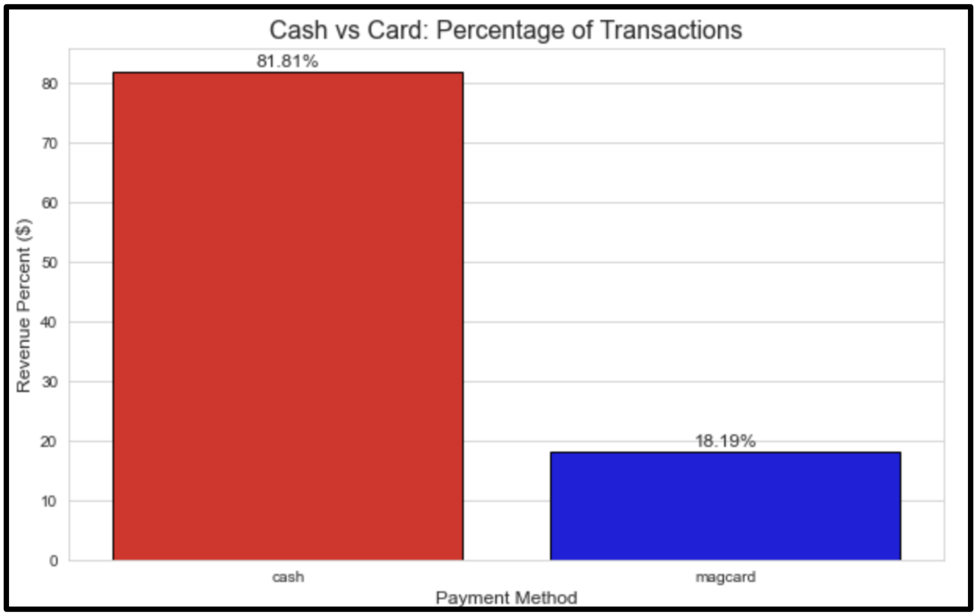

Figure: Percentage of Transactions from Cash and Magcard.

Less than a fifth of payments are made through magcard, while 81% of transactions are done through cash.

We wish to see if the average transaction value is any different between cash and card payments. Perhaps a sizable difference may yield an opportunity. From the graph, the average transaction amount for a cash payment is $8.91 per transaction. However, the average transaction amount in a card payment is over twice as much at $19.65.

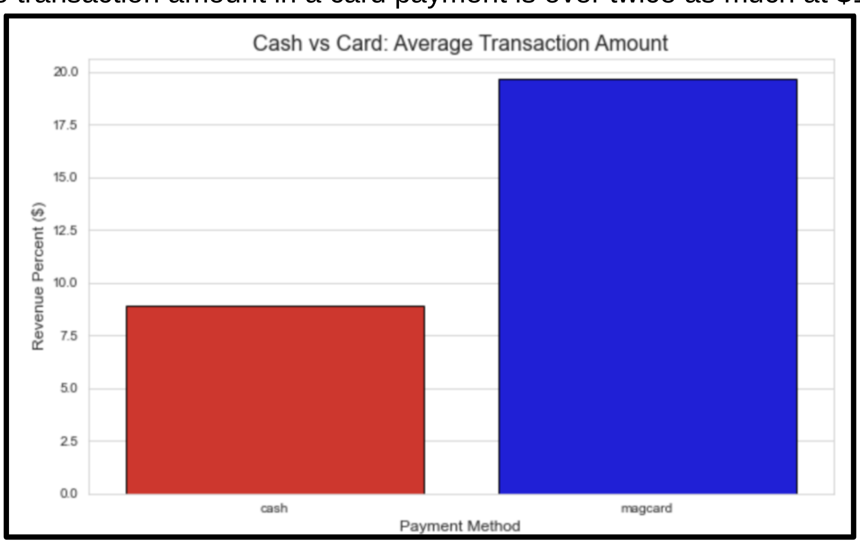

Figure: Average Transaction Amount for Each Payment Method.

Some products are consistently sold at a loss

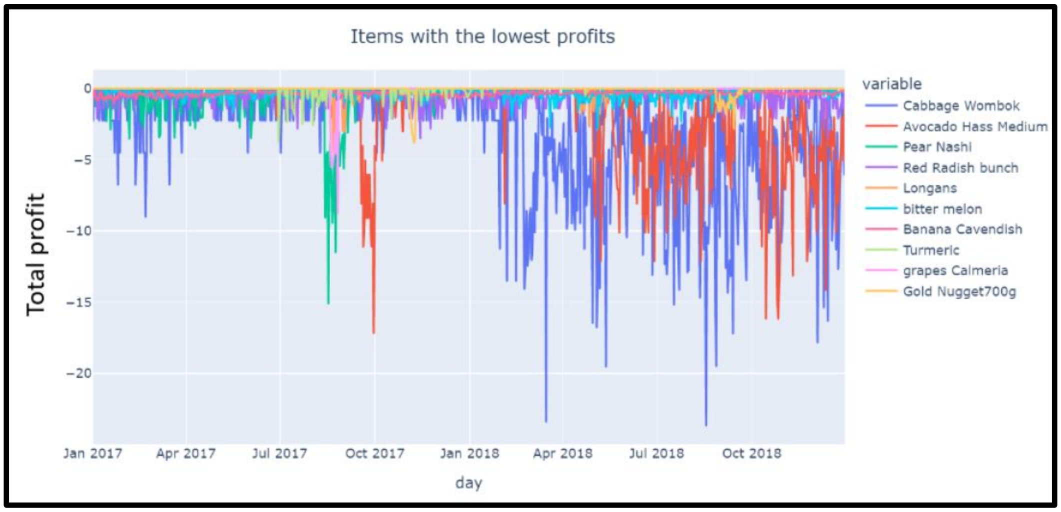

Figure: Shows the lowest profit products over the 2-year span

Some products have great seasonal success

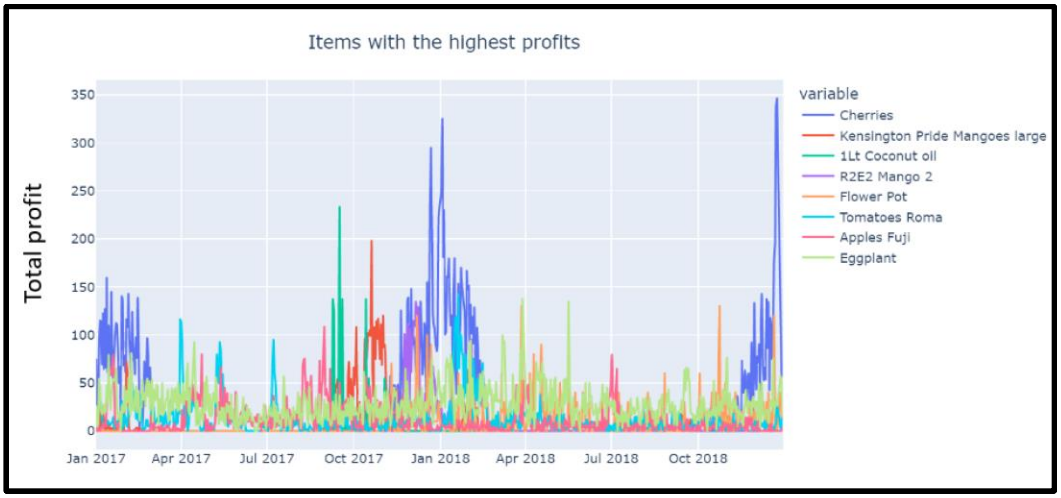

Figure: Shows the highest profit products over the 2-year span

The most profitable months of the year are generally: December, January, February. March. However, December shifted the most from 2017-2018 (-8.87k). Most profitable seasons ranked: Summer, Fall, Spring, Winter.

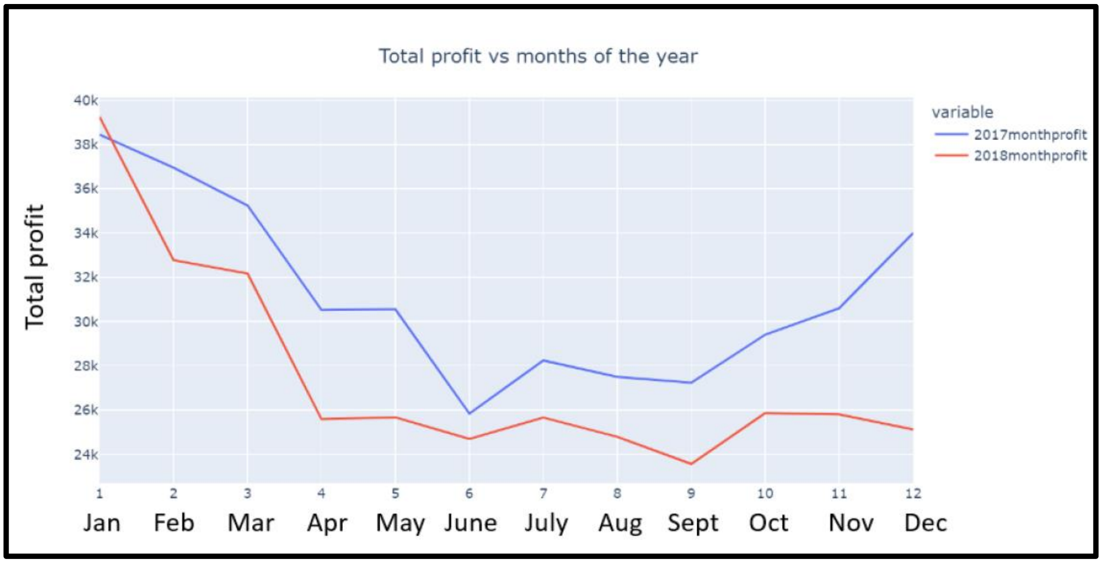

Figure: Shows how profit fluctuates for each month of the year

Saturday is the most profitable day of the week. Sunday is the least profitable day of the week.

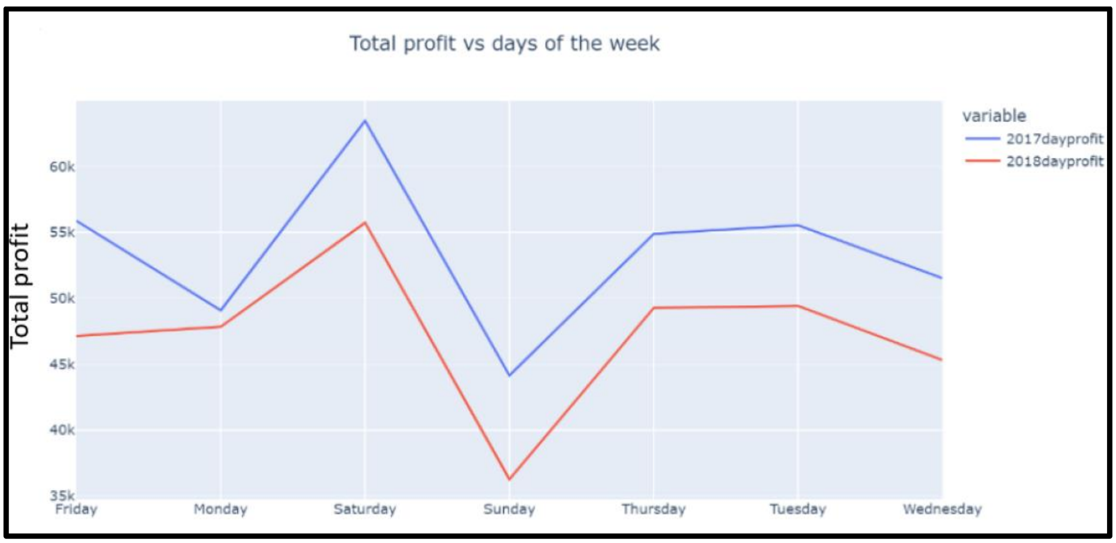

Figure: Shows how profit fluctuates for each day of the week

The most profitable times of day are 10am-12am. The least profitable times of day are 9pm-7am for both. These apply to both years. Additionally, when hours were shortened in 2018, it resulted in an increase in traffic for the remaining hours (as expected), with the biggest increase between 9am and 6pm (This associates with a total profit graph).

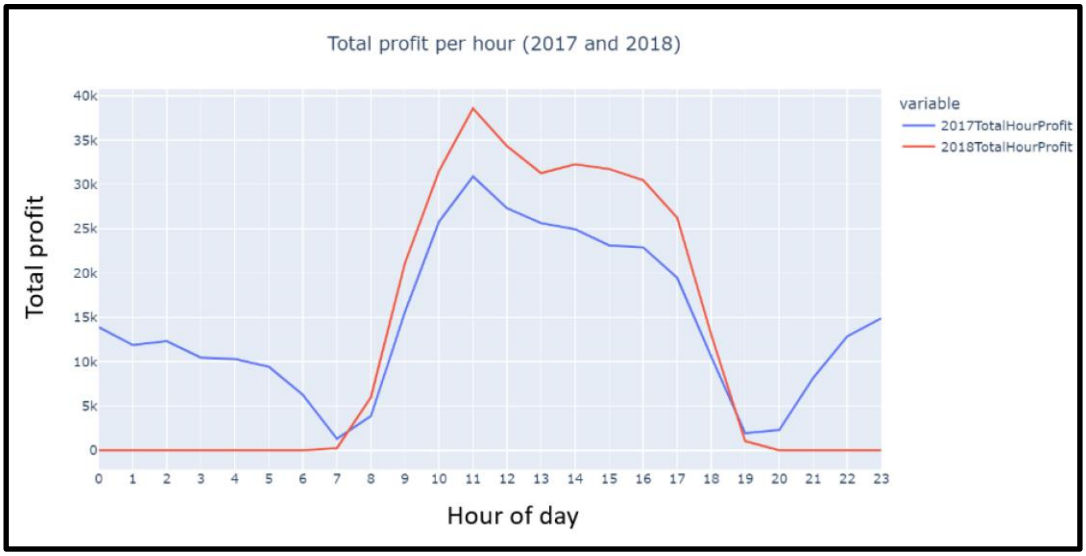

Figure: Shows how the profit per hour fluctuates from both 2017 and 2018

The proportion of profit made by various categories was roughly the same between 2017 and 2018. There were no major shifts.

New Metrics Calculated

We created a new metric, Monthly Profit Margin, that divides the sum of total profits for a sub-category with the total_selling_price and works to understand the percentage of profit made for each subcategory each month.

```python
# Creating margin and volume interactions
for i in range(81):
    val = "High Margin" if mean_var.loc[i, 'Avg. Total_Profit_Margin'] > 0.5 else "Low Margin"
    val += " High Volume" if mean_var.loc[i, 'Quantity'] > mean_var.loc[i, 'Category_Avg_Q'] else " Low Volume"
    mean_var.loc[i, 'MarginRange'] = val

cat_subcat_mar = mean_var[['Main Category', 'Sub Category (group)', 'MarginRange']]
cat_subcat_mar.head(15)
```

	Main Category	Sub Category (group)	MarginRange
0	Bag	Mark down Bag	High Margin Low Volume
1	Fresh Produce	Cut Fruits	Low Margin Low Volume
2	Fresh Produce	Veggies	High Margin Low Volume
3	Fresh Produce	Garlic	High Margin Low Volume
4	Fresh Produce	Chillies	High Margin Low Volume
5	Fresh Produce	Coconut Products	High Margin Low Volume
6	Fresh Produce	Mushrooms	High Margin Low Volume
7	Fresh Produce	Cabbages	Low Margin Low Volume
8	Fresh Produce	Root Vegies	High Margin Low Volume
9	Fresh Produce	capsicum	High Margin Low Volume
10	Fresh Produce	Fruits	High Margin Low Volume
11	Fresh Produce	Other Fruits	High Margin Low Volume
12	Fresh Produce	Avocadoes	Low Margin Low Volume
13	Fresh Produce	carrots	High Margin Low Volume
14	Fresh Produce	Asian Vegies	Low Margin Low Volume

Figure: Code that derives the Margin and Volume Ranges

The next step is tagging every row with the Margin-Volume interaction it falls under. Any sub-category with profit margins over 0.5 is sold at high margin and everything lesser than 0.5 is sold at low margin. If the monthly quantity of the category is more than the average monthly quantity for its main category, it is tagged as high volume; otherwise, it is low volume.

```
def getSubCat(x):
    if x in ['Biscuits', 'Biscuit']: return 'Biscuits'
    elif x in ['Spices', 'spices', 'Spides']: return 'Spices'
    elif x in ['Condiments', 'Condiment']: return 'Condiments'
    elif x in ['Sauces', 'Sauce']: return 'Sauces'
    elif x in ['Packaged', 'Packaed']: return 'Packaged'
    elif x in ['Coconut Products', 'Cocunut Products']: return 'Coconut Products'
    return x

for i in range(le):
    sc = getSubCat(dataa.loc[i,'sub_category'])
    mc = dataa.loc[i,'main_category']

    dataa.loc[i, 'Margin_Vol'] = mean_var[(mean_var['Sub Category (group)'] == sc) &
                                          (mean_var['Main Category'] == mc) ]['MarginRange'].values[0]

dataa.to_csv("cleaned_sales_data_2017_2018 - cleaned_sales_data_2017_2018.csv")
```

Figure: Code that assigns the Margin-Volume Interaction to all of the data

The above helps us understand the strategy for each margin-vol interaction, which will inform our recommendations later on in the document.

Analyzing Pricing Strategies

Comparing the behavior between average monthly margins, average selling price, and average buying prices we could analyze the changes in overall profitability of each sub-category based on its pricing strategy.

We chose cabbages, potatoes, and flowers since they had the most variance in their monthly margins, implying a lot of changes in their pricing strategies.

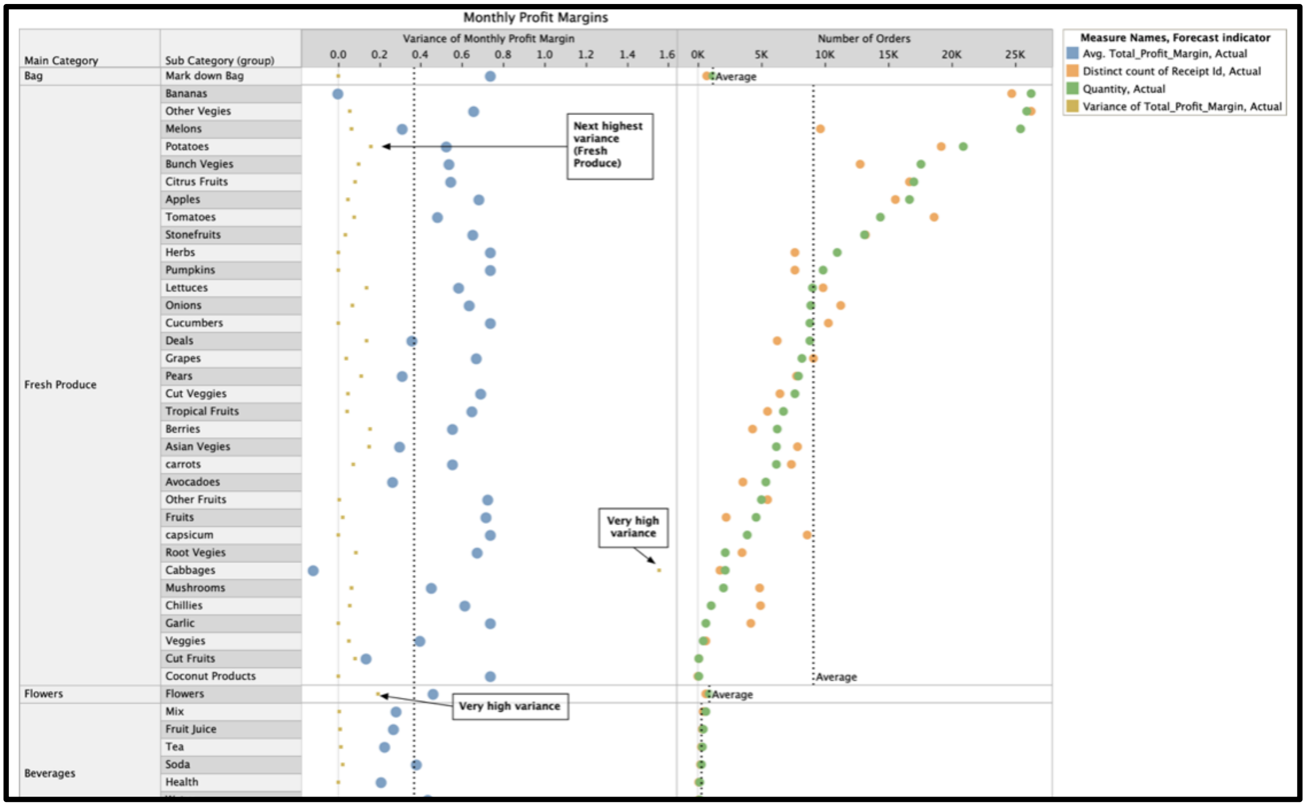

Figure: Graph sorted with the quantity purchased of each subcategory, along with monthly profit margins and their variance

Cabbage was a medium-margin product and its selling price decreased as soon as its buying price increased, ruining the profit margins and turning selling cabbages into a loss. For potatoes, the selling price was adjusted based on the buying price maintaining its profit margins. The best case was for flowers, where a decrease in buying price was exploited by increasing the selling price, improving margins, and turning a low-margin product into a high-margin product.

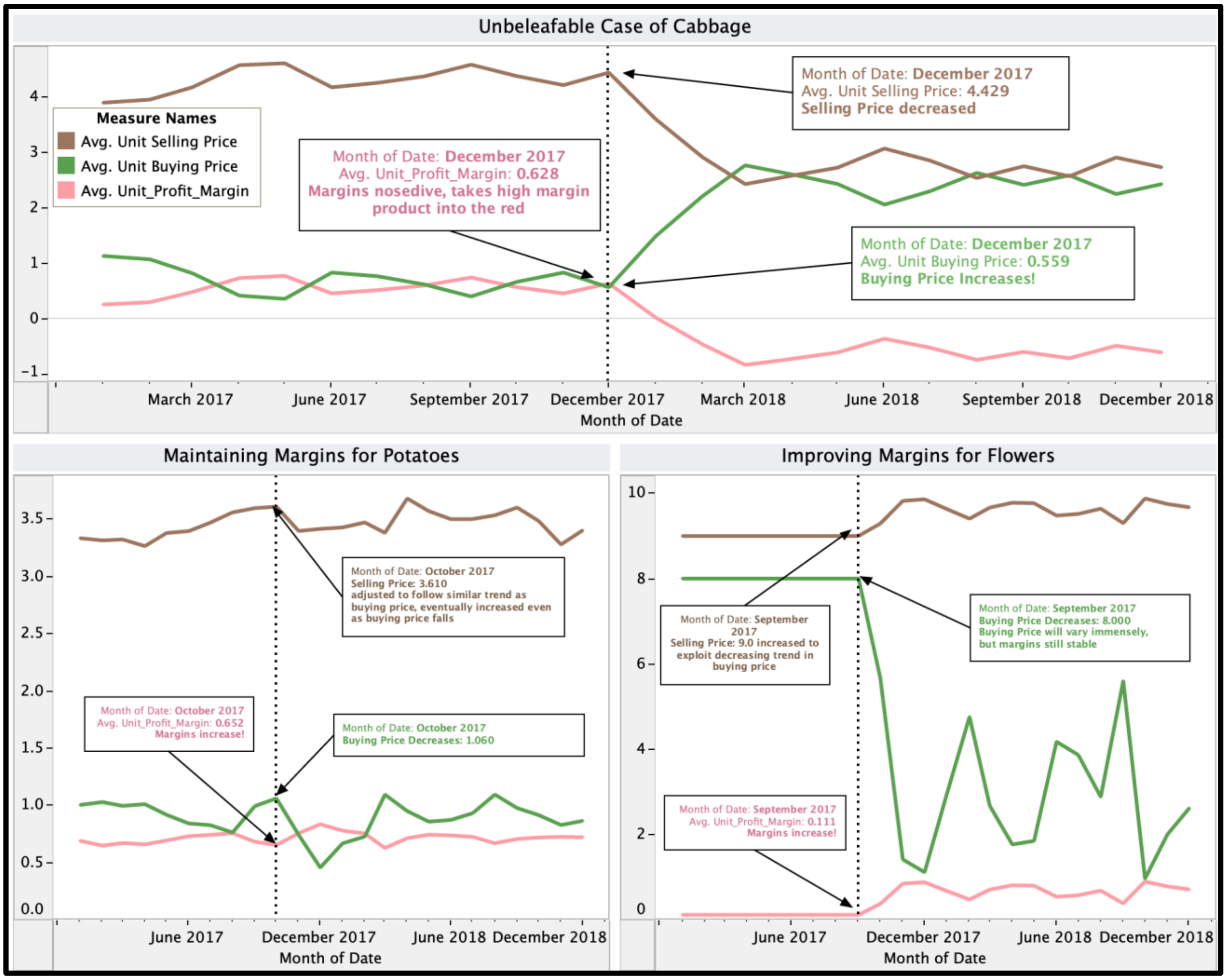

Figure: Graph that demonstrates bad, better, and best pricing strategies

Margin-Volume Interaction

We observed that high-margin and high-volume products are dominated by fresh produce, mostly fruits. Pantry Staples had much lower margins and they could either be low volume (foods purchased less frequently like canned foods, oil, vinegar) or high volume (rice, pasta, spices).

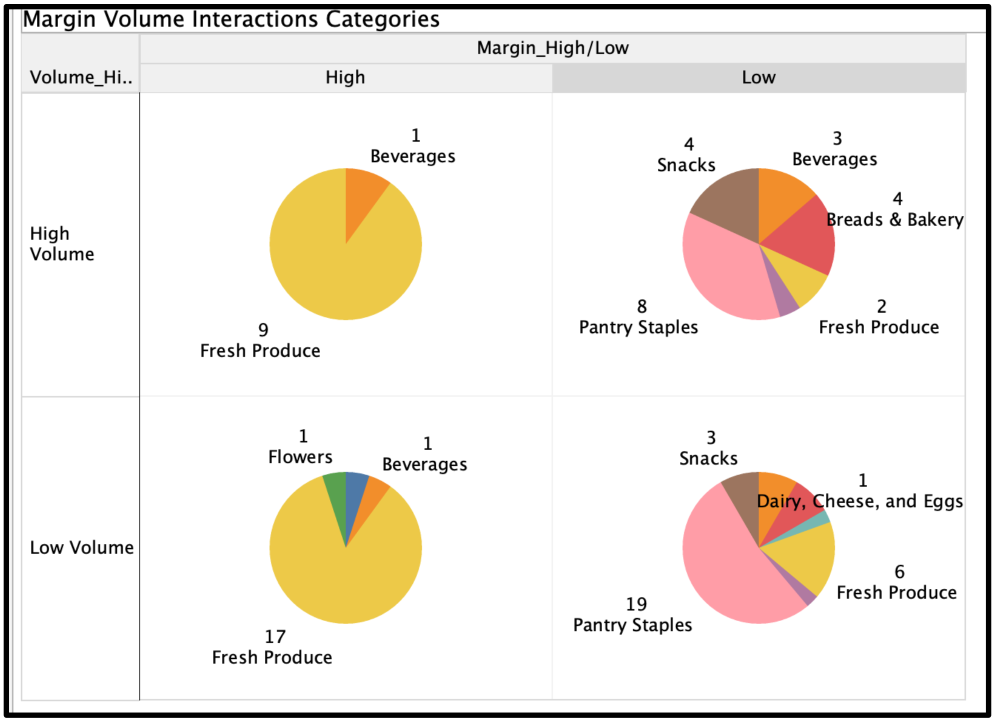

Figure: Graph that demonstrates margin-volume interaction between categories

Fresh produce is mostly sold at a high margin, perhaps due to a lot of them being grown by the retailer. A lot of perishable groceries like dairy, cheese, and eggs were low margin and low volume.

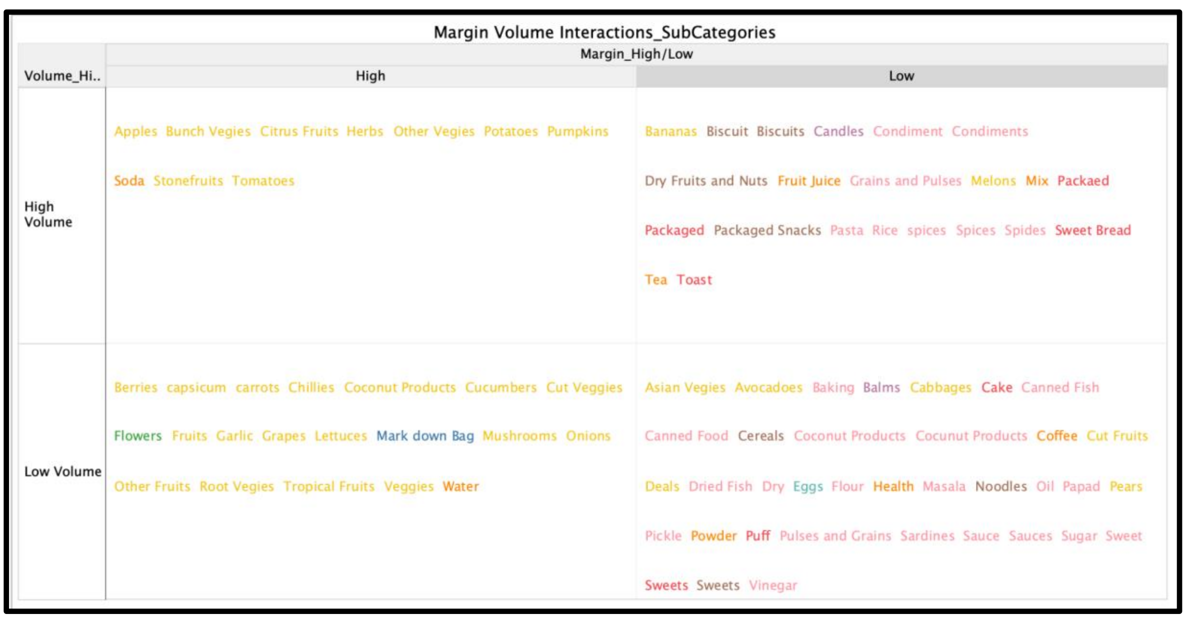

Figure: Graph that demonstrates margin-volume interaction between subcategories

VI. Recommendations:

Based on our analysis, we extend the following recommendations to All Foods:

1. Consider reducing the volume ordered of products that are sold at a loss and are not perishable such as spices like Turmeric, and packaged foods like Gold Nugget 700g (Chicken Nugget).

2. Consider increasing the seasonal volume of products that are sold at a considerable profit: Cherries, Kensington Pride Mangoes Large, 1Lt Coconut Oil, R2E2 Mango 2, Flower Pot, Tomatoes Roma, Apples Fuji, and Eggplant.

3. Increase prices for high volume low margin produce (Cabbages) because there is a lot of profit to be made for products that are already sold a lot, and a better margin would exploit that.

4. For low volume, low-margin products try to increase volume by either offering discounts during the later part of the day (afternoon, as they have the most sales after the 10AM-12PM rush) or on days with high sales (Saturday). A lot of these are perishable and the retailer might be incurring a lot of losses by not selling them off.

5. Some low-margin, high-volume products (bananas have high quantity as well as high number of orders, implying most orders have bananas) are loss leaders and attract people to spend money in the store. Leave the pricing strategy unchanged for products like these that operate at a loss if they are loss leaders.

6. Look into why Sunday is the least profitable day of the week. Consider adding discounts on Sundays to increase volume.

7. December, and specifically the holiday season shows the highest profits. Consider temporarily expanding work hours to 24/7 for December.

8. Look into why colder seasons associate with a lower volume of sales at your store.

9. It would be helpful to place high-volume, low-margin items next to low-volume high-margin products to make the best use of both of the categories. Ex: low-volume produce like onions, mushrooms, and capsicum with pantry staples like pasta, bread, and rice. These would be helpful for people shopping for recipes and help increase volume sold of high-margin products.

10. Maximize card transactions over cash. The average card transaction payment is more than twice the average cash payment, potentially because of a customer's liberal expense patterns with a credit/debit card [4]. Use discounts/cashback as a way to promote the usage of cards. Also, specifically offer marginal discounts on low-volume/high-margin products, which customers might otherwise not buy.

11. Offer appropriate low-volume/high-margin products at a discount when purchased with holiday-related items.

VII. References

1. https://www.abs.gov.au/statistics/economy/national-accounts/australian-system-national-accounts/2018-19#:~:text=The%20Australian%20economy%20expanded%20by,1.9%25%20in%20the%20previous%20year
2. **Git Repo:** https://github.com/logan-obrien/DataThon_Data_Brokers.git
3. **Tableau Public:** https://public.tableau.com/app/profile/arjun.sharma8097/viz/DataBrokersAllFoodAnalysis_16768435147820/Story1?publish=yes
4. **You are more likely to overspend on a credit card over cash:** https://www.nerdwallet.com/article/credit-cards/credit-cards-make-you-spend-more#:~:text=For%20starters%2C%20thanks%20to%20federal,as%20extended%20warranties%20and%20insurance