Grocery Shopping Behavior Analysis

Group 6: Fangzhou Xiang
Jinman Rong
Ziyu Tang

Overview of trips, households, and stores

7,596,145
Shopping trips

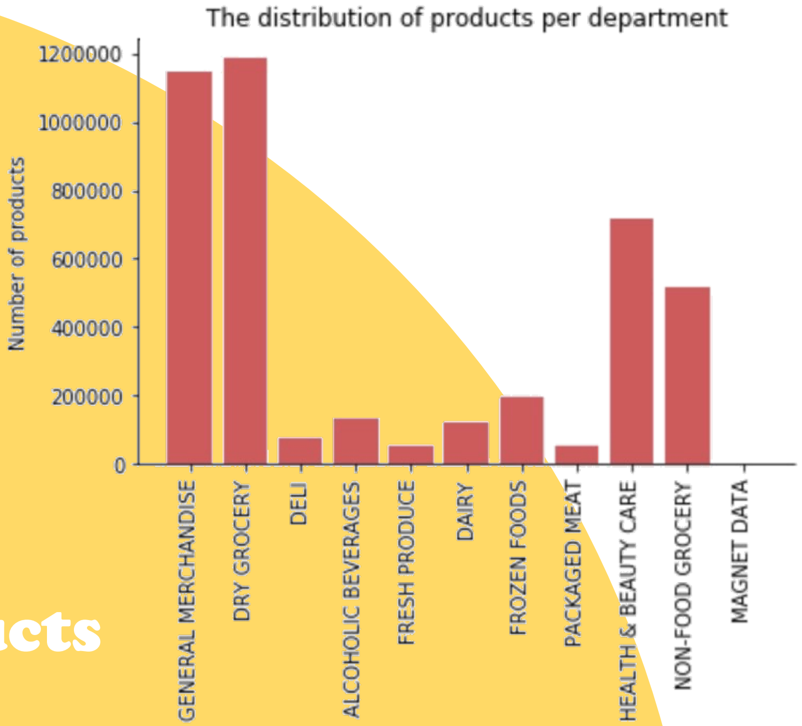
39,577
Households

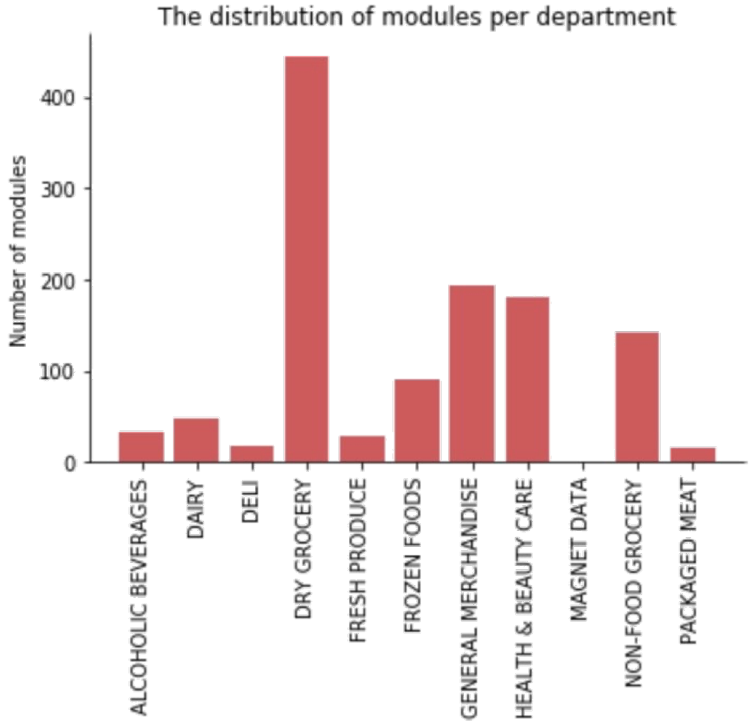
26,406
Stores of different retailers

Overview of products

4,231,283

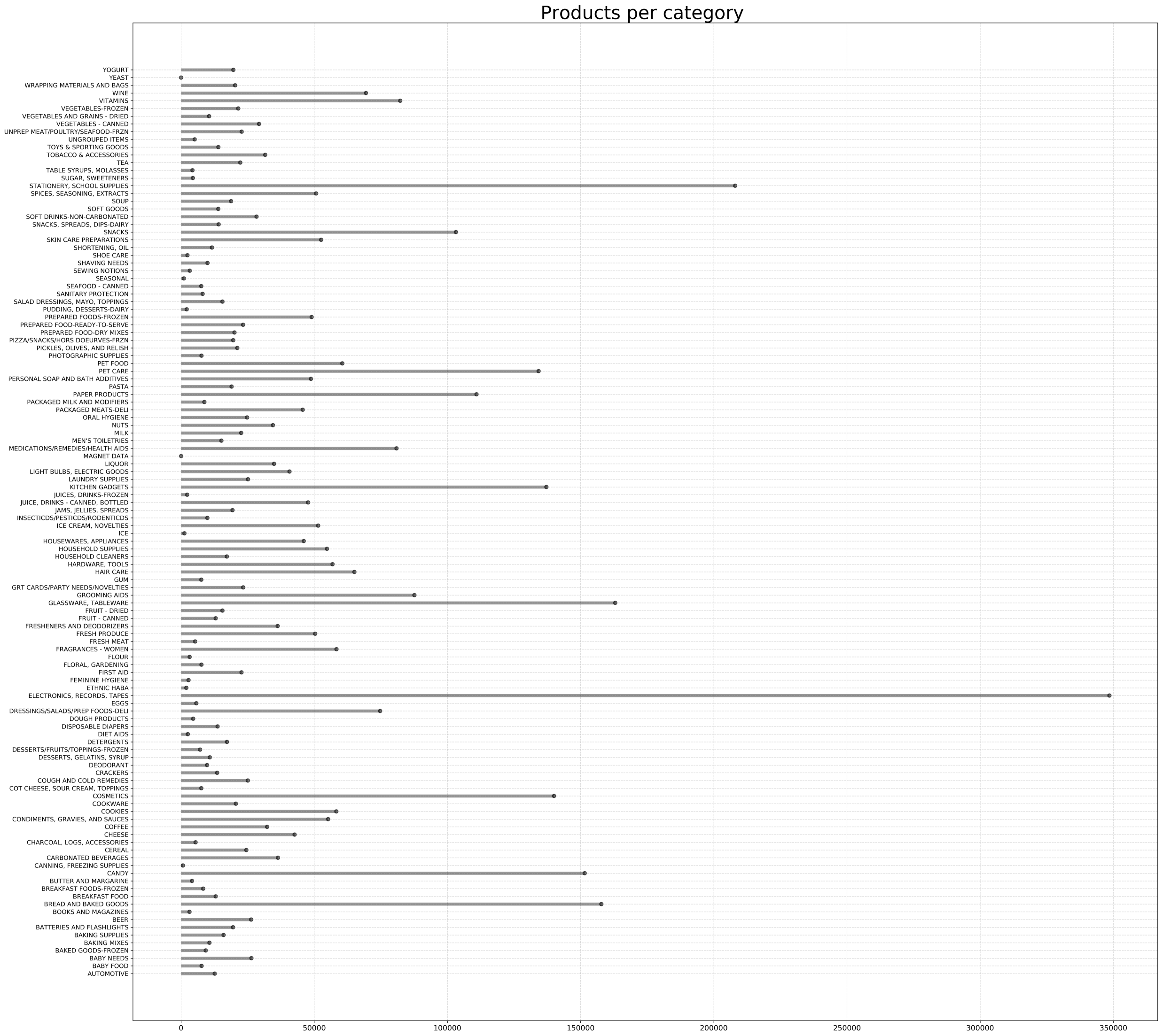
Different products





Dry grocery has the most of products and modules among the departments.

Overview of products

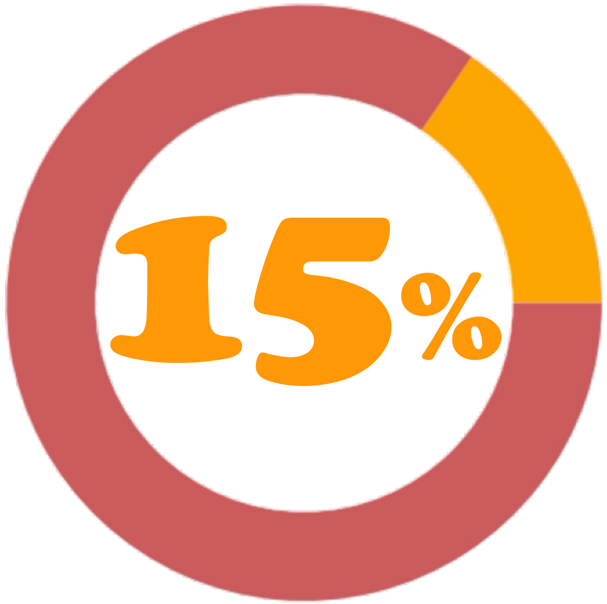


Different products per category

Remarks:

- Products per module has 1224 records, you can see detailed information in the table “product per module”.

Overview of transactions



7,596,145 Total transactions

Data from table “Trips”

Transactions with coupon

874,873

Data from table “Purchases”

Remarks: The total transactions in table “Trips” is 7,596,145, while the one in table “Purchases” is 5,651,255.
We use data of total transactions from table “Trips”, and data of transactions under promotion as well as its proportion from in table “Purchases”.

Customer Analysis

48 households don’t shop at least once on a 3-month period.

We set the start date as 2003-12-27, one day before the earliest record from the data, and the end date as 2004-12-26, one day after the latest record from the data. The inactive customers are defined if time window between two consecutive shopping trips is more than 90 days during this one-year period. We use metric “90” for the reason that the minimum length of 3 months is 90 days in 2004. Finally, 48 household are found as inactive customers.

Is it reasonable?

It’s not reasonable for a household do not shop on a 3-month period or more, because people always have something to buy, like food. Something must happen to these households.

What happens?

- Go for holidays
- New Comer
- Leaver

Customer Analysis

Reasons for no shopping

Go for holidays

Shopping record is interrupted because of holidays.

Typical customer: 2012007

Characteristic:
Only have no shop behavior from June to August, which can be a summer vacation.

New Comer

Shopping record first occurs in March or later.

Typical customer: 8321524

Characteristic:
Only start to shop from March in the 2004, indicating that they may be the new comers to the community .

Leaver

Shopping record last occurs in September or before.

Typical customer: 9164546

Characteristic:
Stop shopping from September, indicating that they may leave the community.

Loyalism

35,962 Households shop at least once a month

0.34% : "Crazy Fans"

Of them concentrate at least 80% of their grocery expenditure on **single** retailer.

0.88% : "Loyal Members"

Of them concentrate at least 80% of their grocery expenditure among **two** retailer.

Remarks:
- 124 households concentrate at least 80% of their grocery expenditure on single retailer in 12 months. And 316 among 2 retailers.
- We use grocery expenditure month by month instead of average monthly expenditure to get more accurate results.

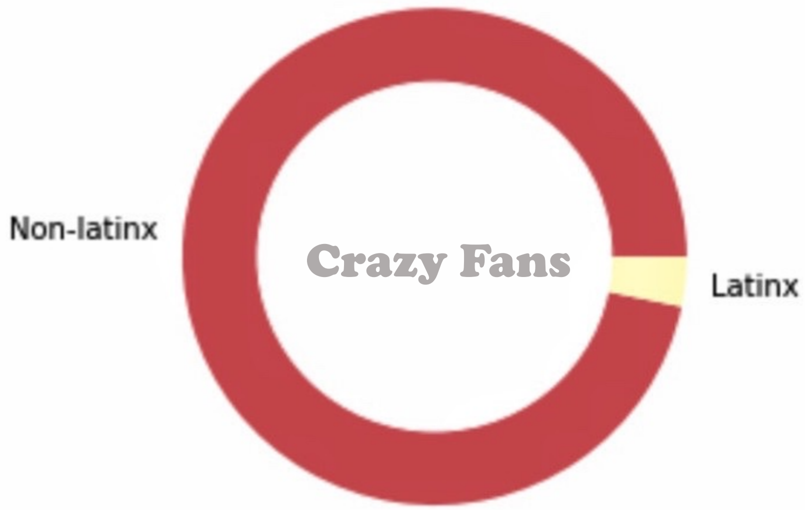
Loyalism

Difference in race

Most of crazy fans and loyal members are White Caucasian, while Asian is always the least. The proportion of African American in loyal members becomes larger than it in crazy fans.

Loyalism

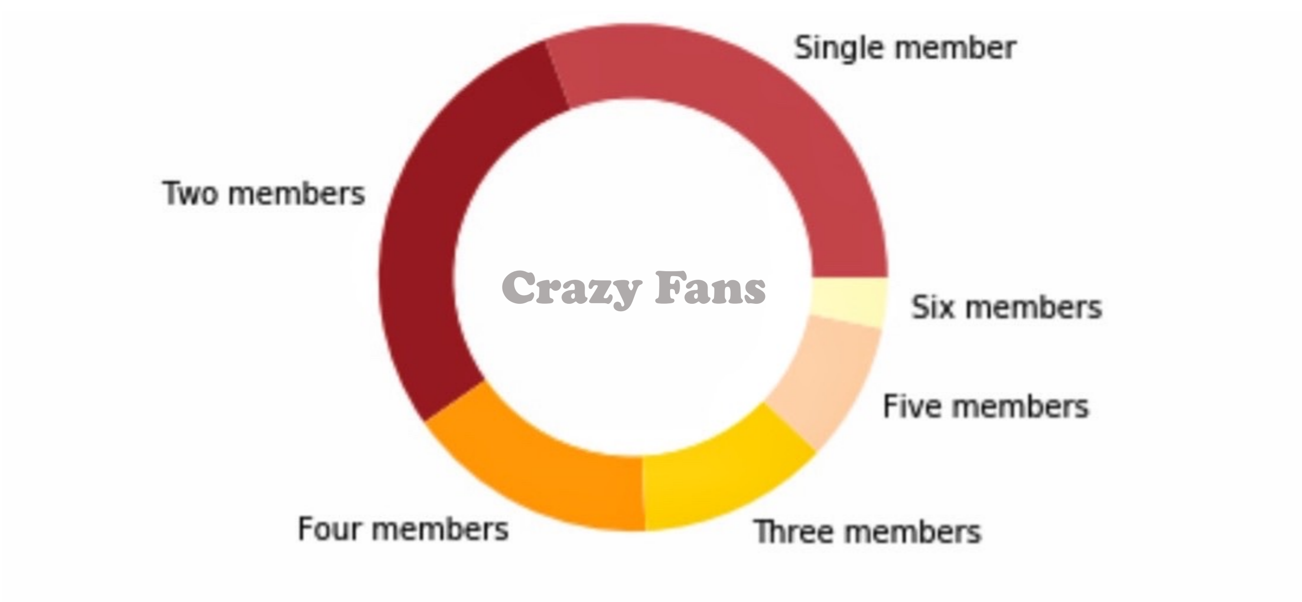
Difference in Latinx

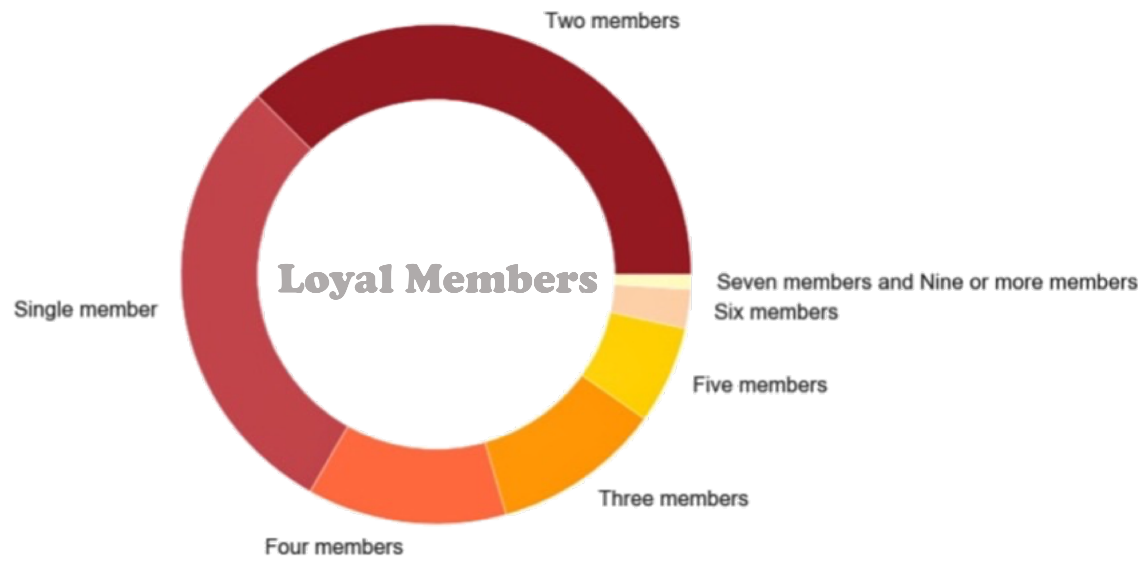


Non-latinx always takes up the most proportion of crazy fans and loyal members.

Loyalism

Difference in size

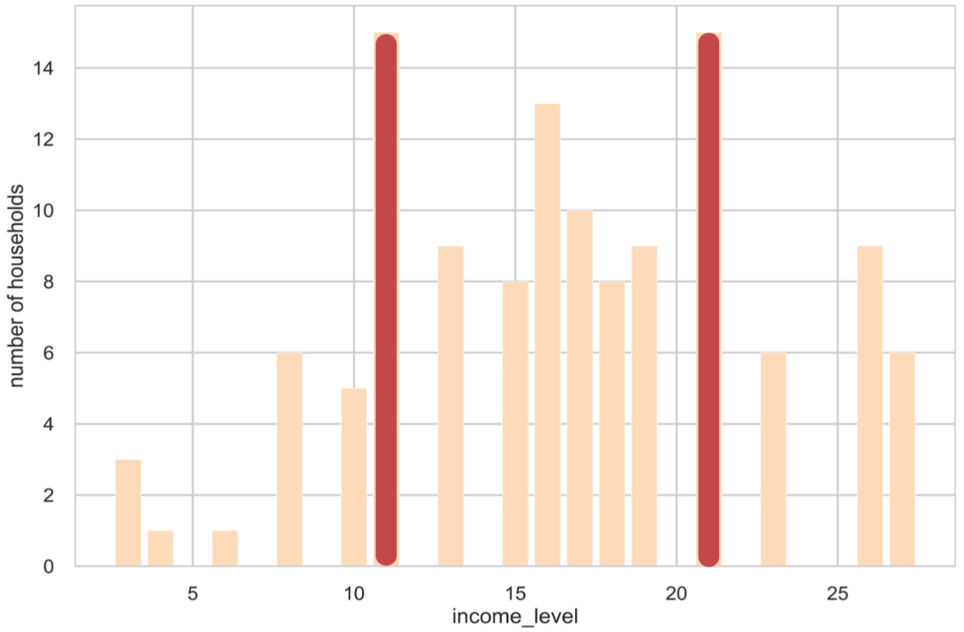




Single-member and two-member family are always top 2 of crazy fans and loyal members.
Two-member family becomes top 1 in the loyal members, indicating that two people may have different tastes about retailers to go. Maybe they always go to the retailers near their workplaces so that the family, as the whole, has two consistent retailers to purchase.

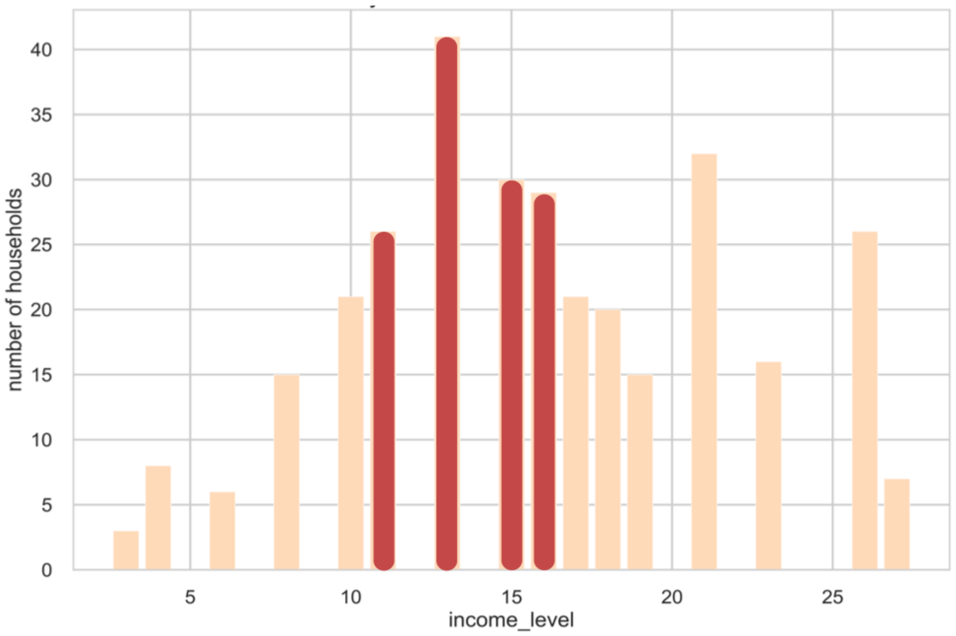
Loyalism

Difference in income



Crazy Fans

Crazy fans' income distribution is approximately M curve: lower middle income and upper middle income are the most frequent observation.

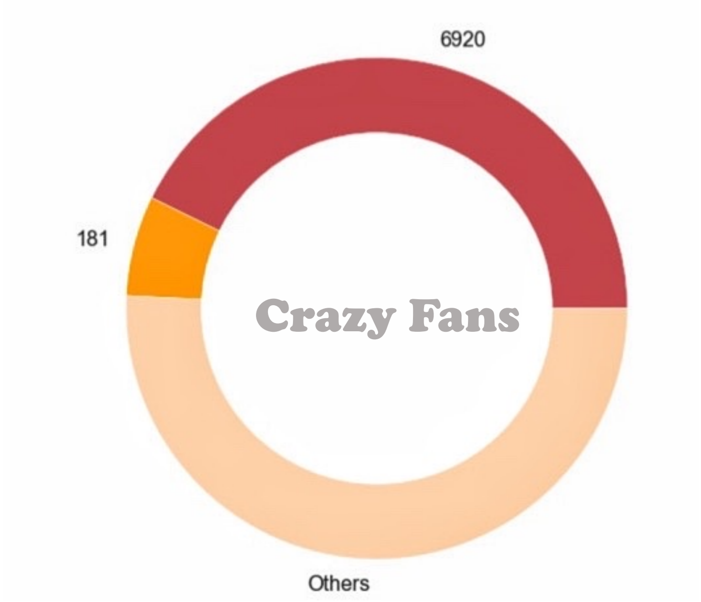


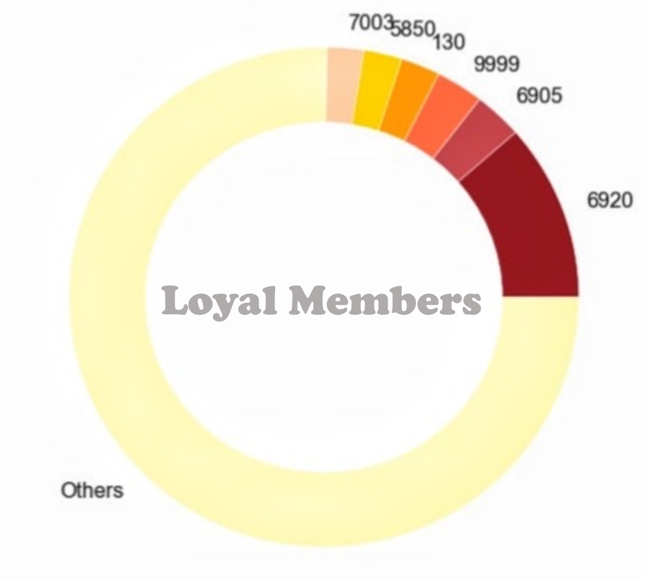
Loyal Members

Loyal members' income distribution concentrates to lower middle income. Household with about 25k yearly income is more likely to be observed.

Loyalism

Difference in retailers



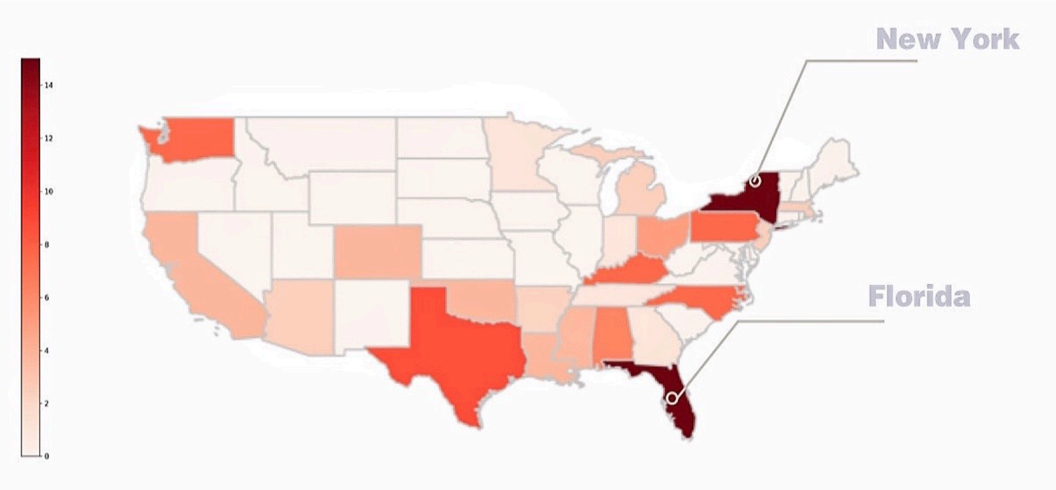


For the crazy fans' market, 6920 and 181 are the top 2 biggest retailers, taking up nearly half of the market. Besides, 6920 is the absolute leader.

Things are different for loyal members' market. 6920 is still the leader of the top 6 retailers but doesn't occupy the majority of the market. Small retailers make up a large proportion, showing the long tail effect.

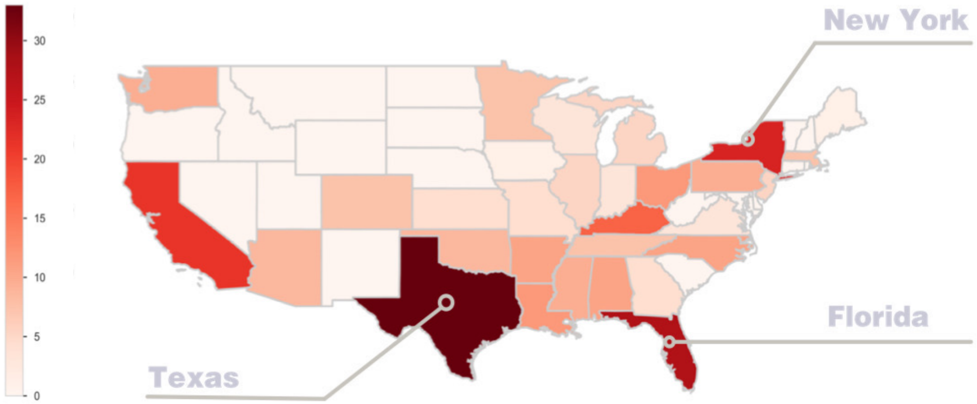
Loyalism

Difference in distribution by state



Crazy Fans

The crazy fans' markets mostly distribute in New York and Florida.



Loyal Members

For the loyal members' markets, Texas comes into being and even become the top 1. New York and Florida are still the top states with most loyal members.

Loyalism

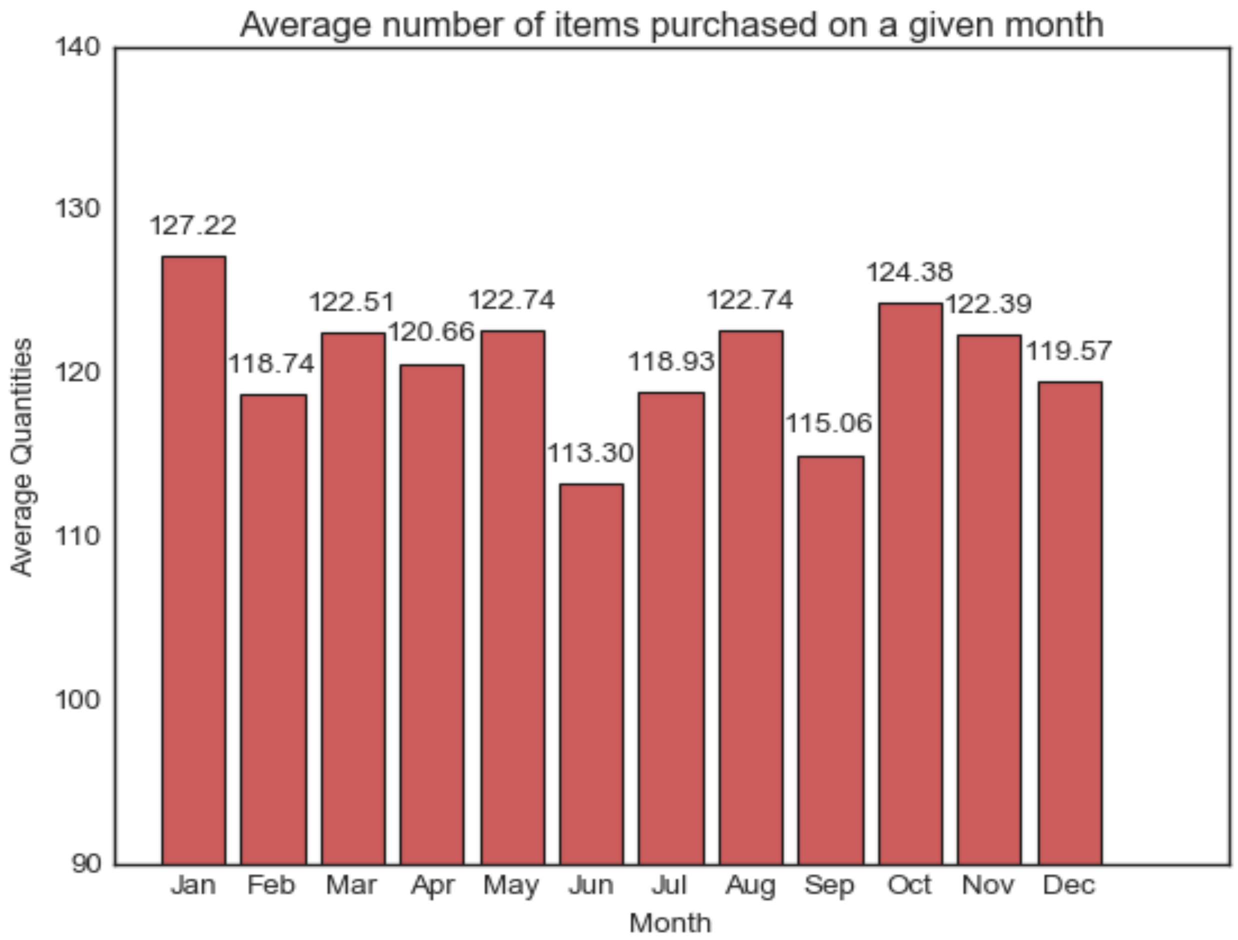
Difference in residence type

Condo

It’s worth noting that there are nearly no crazy fans or loyal customers in one family house-condo or two family house-condo. It may because that these kinds of residence types are often used for rent, such as Airbnb. Seldom do long-term residents live in such kinds of houses, resulting in less stable and loyal shopping behavior observed.

Customer Analysis

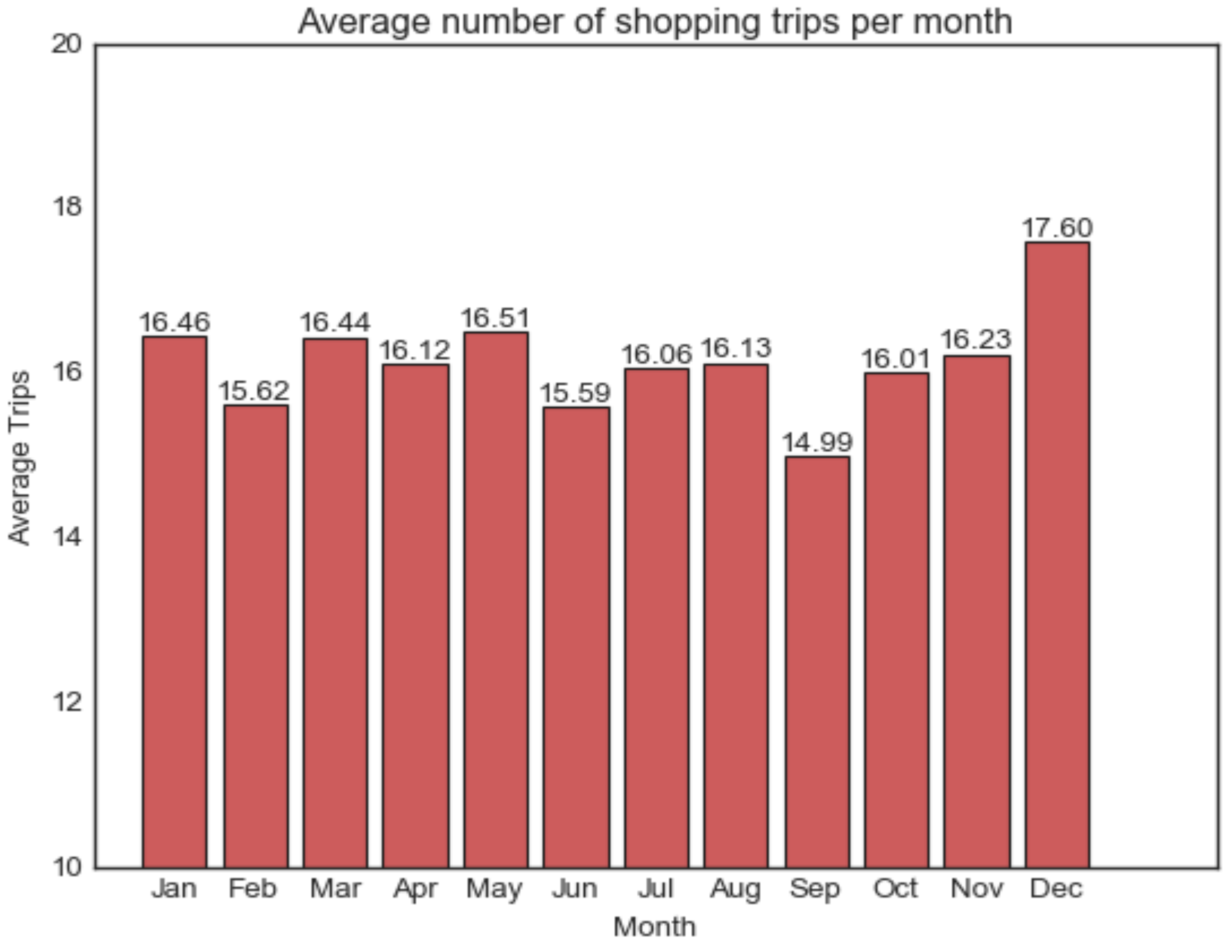
Average number of items purchased on a given month



January contains the most average number of items purchased, while June has the least.
It can be seen that people purchase more around Christmas and Thanksgiving day.

Customer Analysis

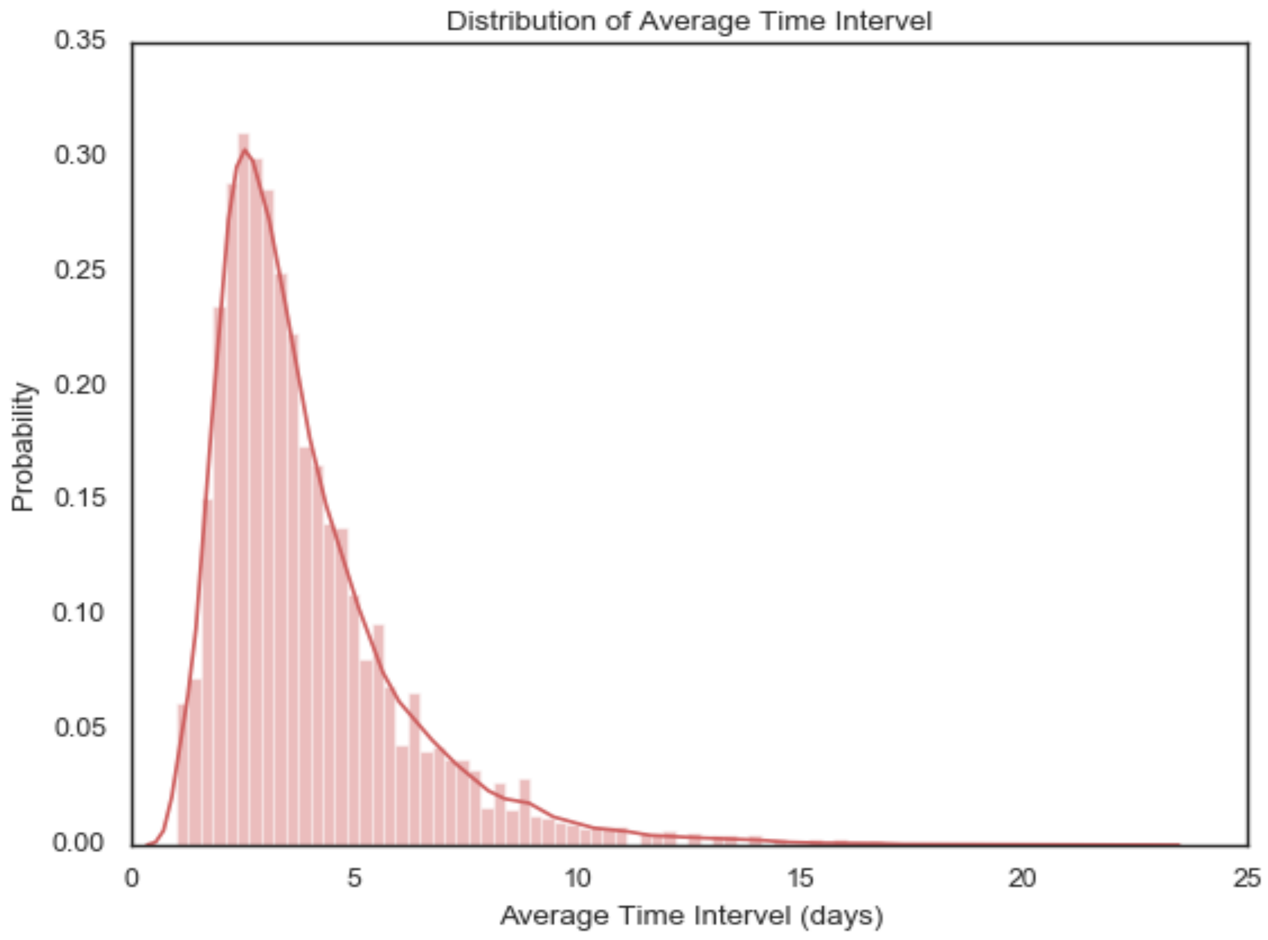
Average number of shopping trips per month



December experiences the most average number of trips per month, while September has the least.

Customer Analysis

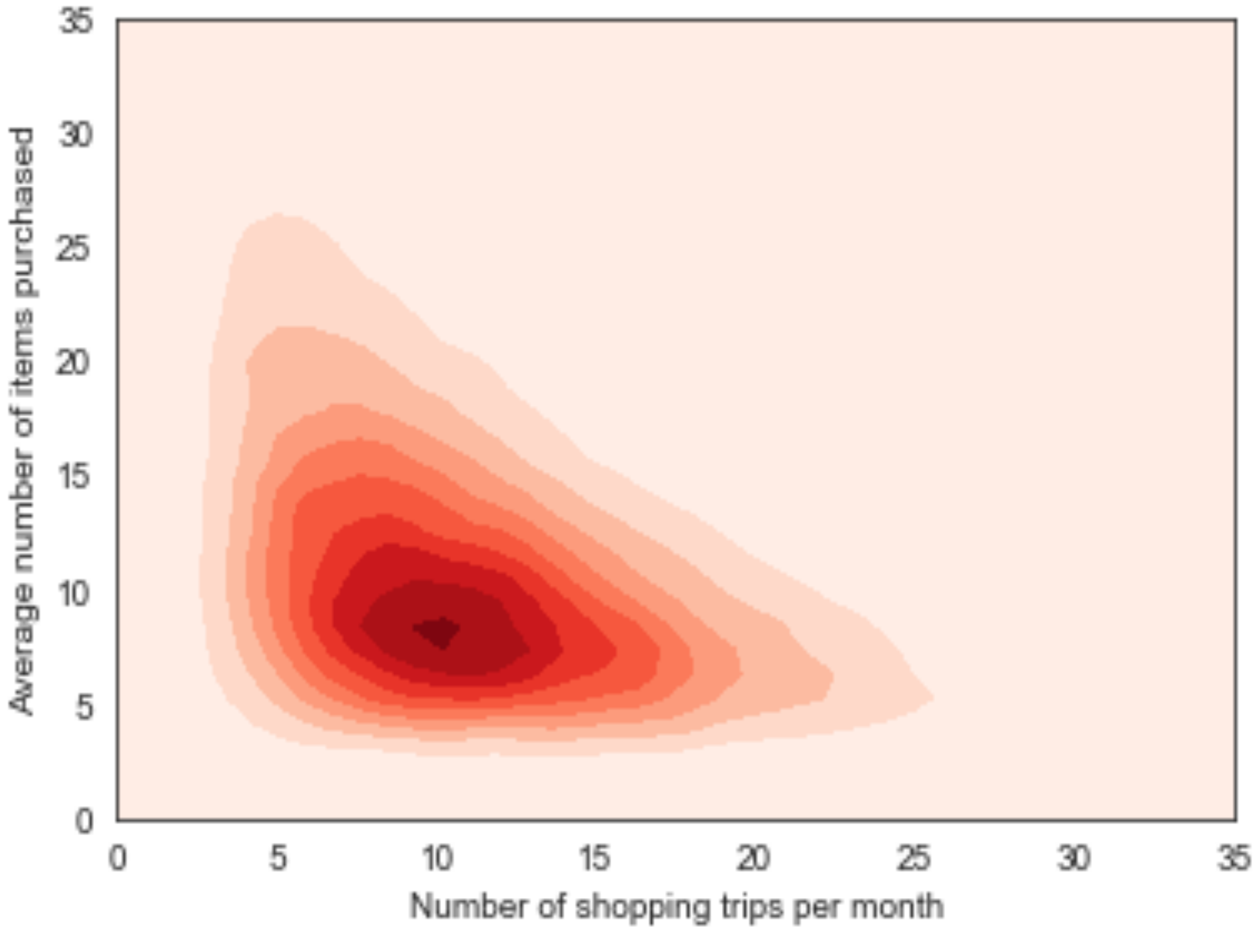
Average intervals between 2 consecutive shopping trips



Most people’s shopping time interval is about two days. The larger the interval, the fewer the observations. And few people have shopping intervals of half a month or more.

Customer Analysis

Relationship between shopping trips and items



The number of shopping trips per month is negatively correlated with the average number of items purchased.

It's because that if people live near the shops and purchase frequently, they don't need to buy lots of items at once. The less frequently people shop, the more items they need to buy at once. Shopping about 10 times a month, 8 items at once is the most common purchase behavior. Purchasing every day but few items at once, or seldom purchasing but buying lots of items at once are the most rare purchase behaviors.

Customer Analysis

Relationship between price and purchase quantity

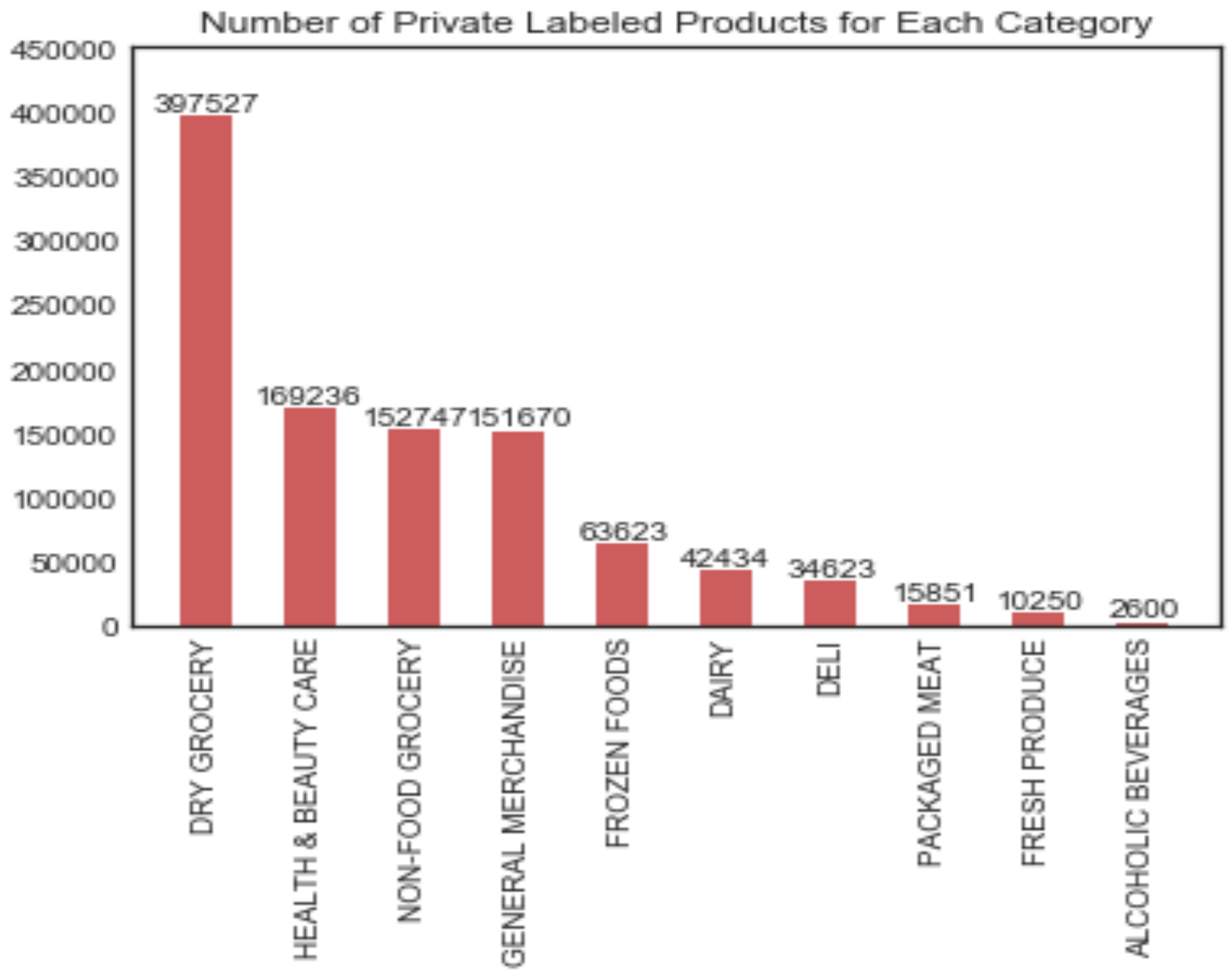
The average price paid per item is negatively correlated with the number of items purchased.

It's because that the items with high daily demands are usually with comparatively low prices, such as bottled water. By contrast, the purchase quantities of high price items, such as electronic products, are comparatively less.

Besides, outliers are seldom observed, indicating that people don't buy too much items even though they're cheap. They purchase on demand, not just price.

Private Labeled products

Categories with “Private labelled” products

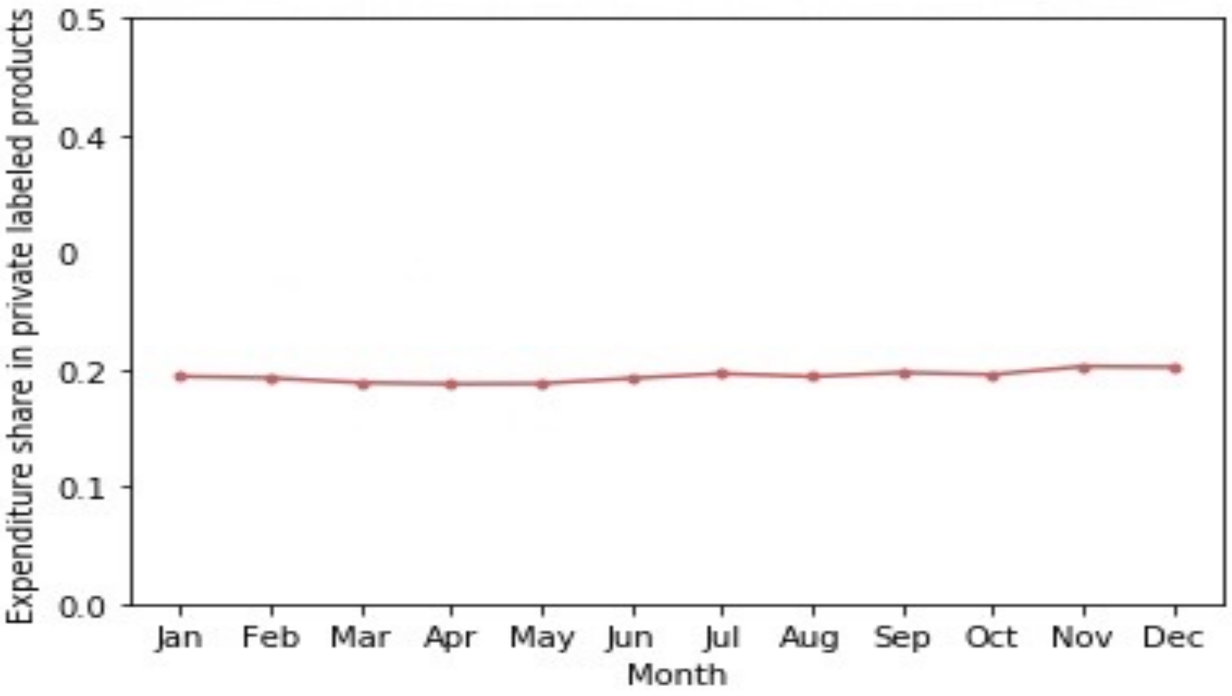


The dry grocery and health& beauty care are the top 2 categories that have proven to be more “Private labelled”.

The categories with huge demands are suitable for establishing private brands. Retailers can build their own brands in these categories and adopt the cost leading strategy, leveraging economies of scale to make profit.

Private Labeled products

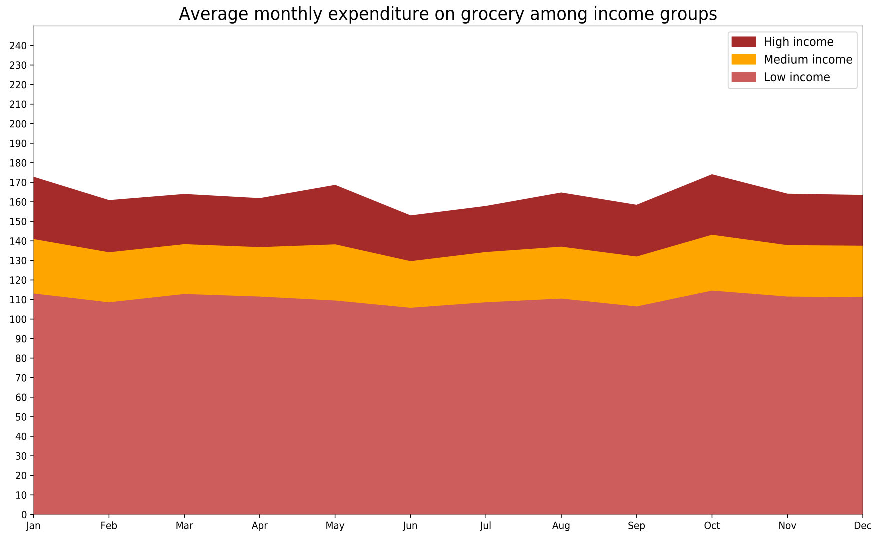
Trend of expenditure share in private labeled products

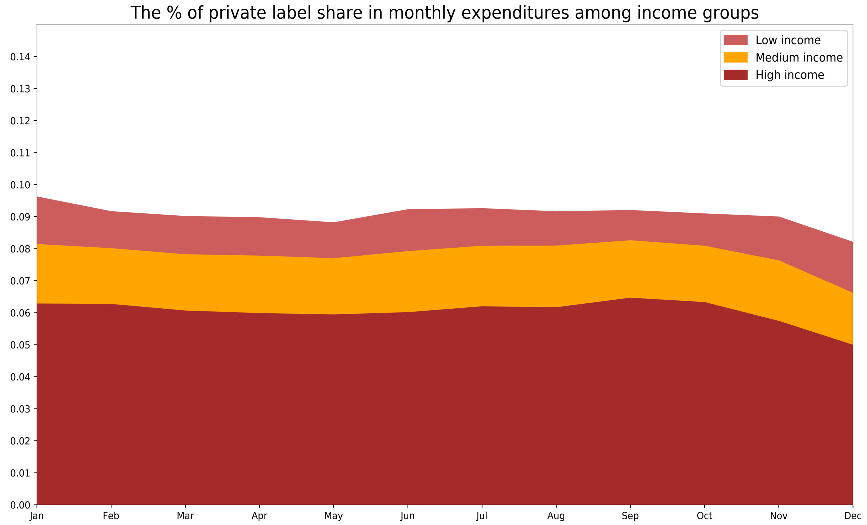


The expenditure share in Private Labeled products is constant across months. It is because that people's shopping behavior is relatively stable, and shopping preferences will not change much across the months.

Private Labeled products

Different purchase behaviors among 3 income groups





High income groups have the most average monthly expenditure on grocery, while low income groups have the least, indicating that high income groups have more budget to buy groceries and purchase more.
Interestingly, the percentage of private label share in monthly expenditure experiences the opposite situation. Low income groups have the most share of private label in monthly expenditure while high income groups have the least. It is because that private label is usually with low price which meets the purchase preference of low income groups . For the high income groups, however, they are more concerned about the quality of goods and care less about the price. As a result, they may buy more expensive, non-private labeled goods than cheaper, private labeled ones.

Thanks for watching

Group 6: Fangzhou Xiang
Jinman Rong
Ziyu Tang