



DSA3101 Assignment 1

1 September 2020



Meet the team!



Chan Yu Hang



Chua Cheng Ling



Chua Huaren



Cai An Qi

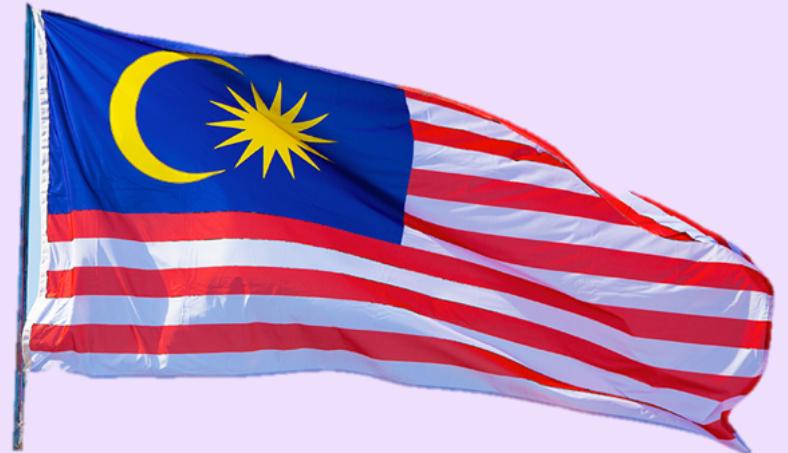


Joash Chin



Clarence Ong

Problem Statement



Country

Retail trade industry situated in Malaysia.

Dataset

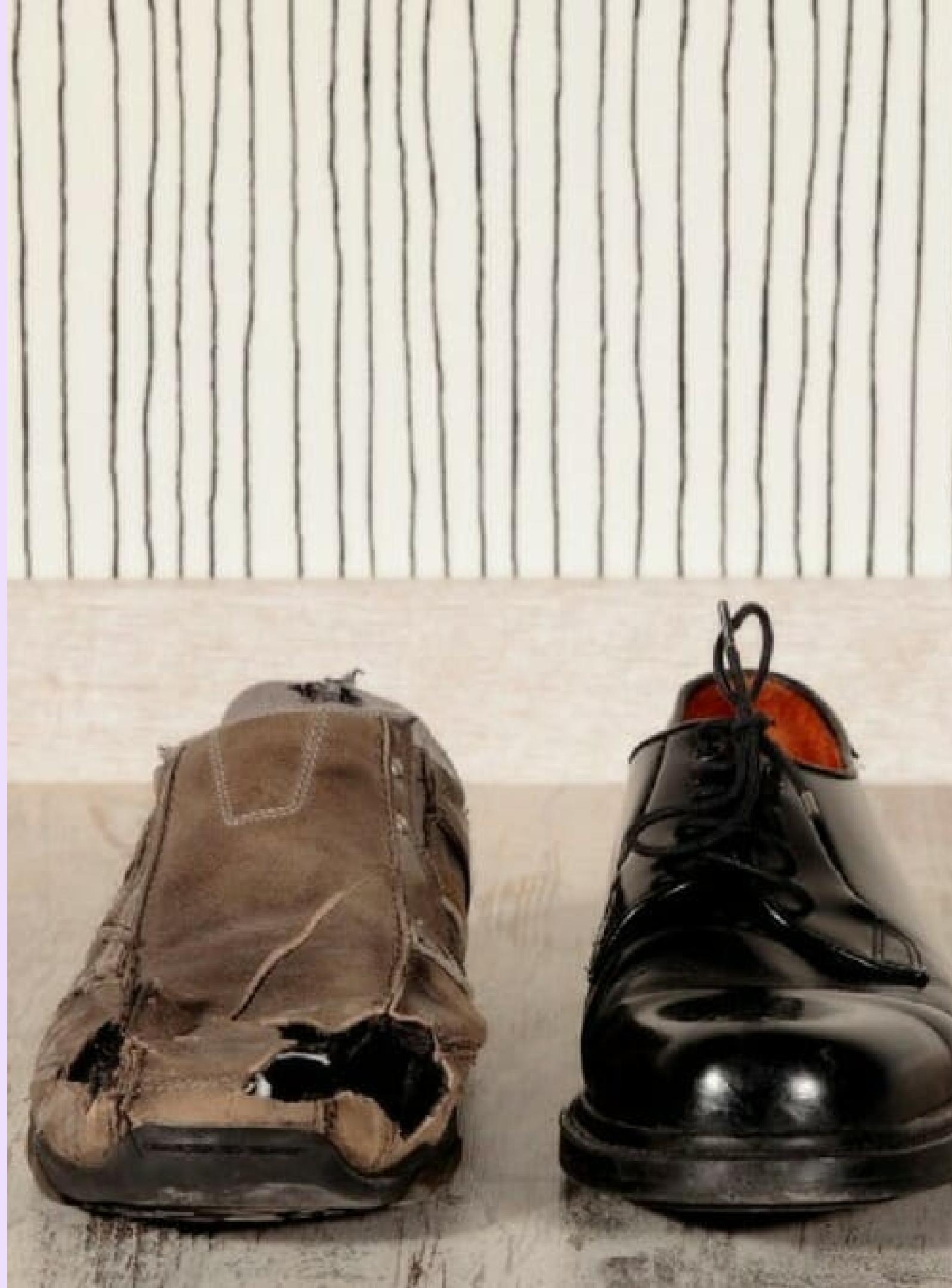
2017-06-25 to 2020-06-14
62 categories of retail items

101011101	2017-10-22	Milk Powder-Adult	1	0.7	29.6
101011101	2017-10-22	Coffee	1	0.3	15.4
101011101	2017-11-12	Biscuits	10	1.0	16.0
101011101	2017-11-12	Drinking Water	3	36.0	20.7
101011101	2017-11-19	Coffee	1	0.1	2.0
101011101	2017-11-19	Cordials	1	2.0	8.9
101011101	2017-11-26	Margarine	1	0.2	5.2



Questions

How can we maximise revenue for the client?
How can we improve brand loyalty of the customers?



Pre-Processing

Group the different panels by their income level and as such label them as categorical socio-economic status.

Various SES levels that we have are:

[0, 1500), [1500, 2000), [2000, 3000), [3000, 4000),
[4000, 5000) , [5000,)

Dealing with "erroneous" data

Removing data

For "erroneous" data, we just remove those rows. These rows make up < 5% of the entire dataset.

Regression imputation with scikit learn

Imputing erroneous cells with mean/median



Malaysia's Movement Control Order



Income difference

Central (Capital of Malaysia - Kuala Lumpur) and South (Closer to Singapore) regions are generally more wealthy than North and East Coast.

location	[0, 1500)	[1500, 2000)	[2000, 3000)	[3000, 4000)	[4000, 5000)	[5000,)
Central	0.0477164	0.0743013	0.1806408	0.1792774	0.1199727	0.3980913
East Coast	0.2708601	0.1450578	0.2169448	0.1270860	0.0770218	0.1630295
North	0.1981651	0.1422018	0.2201835	0.1605505	0.0871560	0.1917431
South	0.1202899	0.1144928	0.2173913	0.1971014	0.1318841	0.2188406

Recommendation

Increased pricing options for central and southern markets.

RFM MODELLING

Assumptions

01

Weekly data represents the purchases a customer has made in that particular week.

Since there is no data on receipts, we shall treat this as a single visit for computing frequency score.

Approach

02

Group by the Panel ID and Date. Then group by Panel ID to apply RFM.

The monetary is computed based on mean expenditure as the total sum is easily influenced by the frequency of the visits (collinearity issues)

R

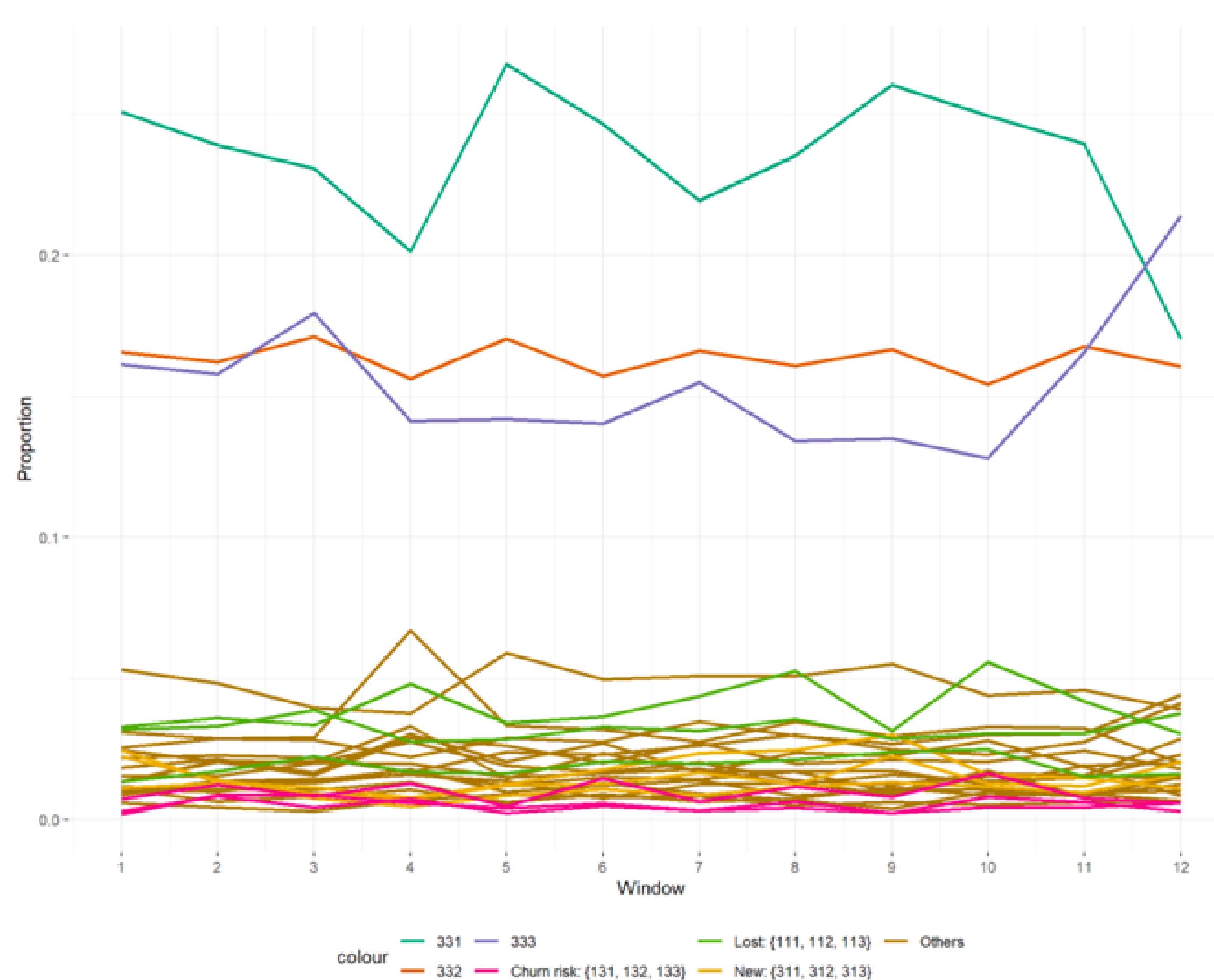
- 3: Last visit \leq 1 week ago
- 2: Last visit \leq 2 weeks ago
- 1: Last visit $>$ 2 weeks ago

F

- 3: \geq 9 visits in 12 weeks
- 2: < 9 visits in 12 weeks
- 1: < 6 visits in 12 weeks

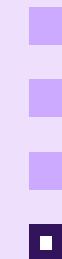
M

- 3: \geq RM60 in a week
- 2: < RM60 in a week
- 1: < RM30 in a week



Consistent loyal customers

Under the MCO, the proportion of high-value customers with M=3 has increased.



Pre MCO VS Post MCO Period

Change	n	Proportion
No Change	560	0.2819738
332 -> 333	123	0.0619335
331 -> 332	107	0.0538771
331 -> 333	40	0.0201410
332 -> 331	40	0.0201410
333 -> 332	39	0.0196375

333 Customers

Many customers are spending more during the pandemic.

Increase in customers with larger M values

Income	Counts	Proportion
[0, 1500)	88	0.1749503
[1500, 2000)	64	0.1272366
[2000, 3000)	102	0.2027833
[3000, 4000)	76	0.1510934
[4000, 5000)	59	0.1172962
[5000,)	114	0.2266402

331 Customers

For the customers with high R and F, a significant number of them had low M. Amongst these 331 customers, most of them have high income.

Recommendation

Target more "331" customers and promote more ads for
‘luxury items’ that cost more.

MBA ANALYSIS

01

Observations

Among loyal customers in the high income bracket, whilst most are valuable customers (RFM = 333), there is still a sizeable group with RFM = 331

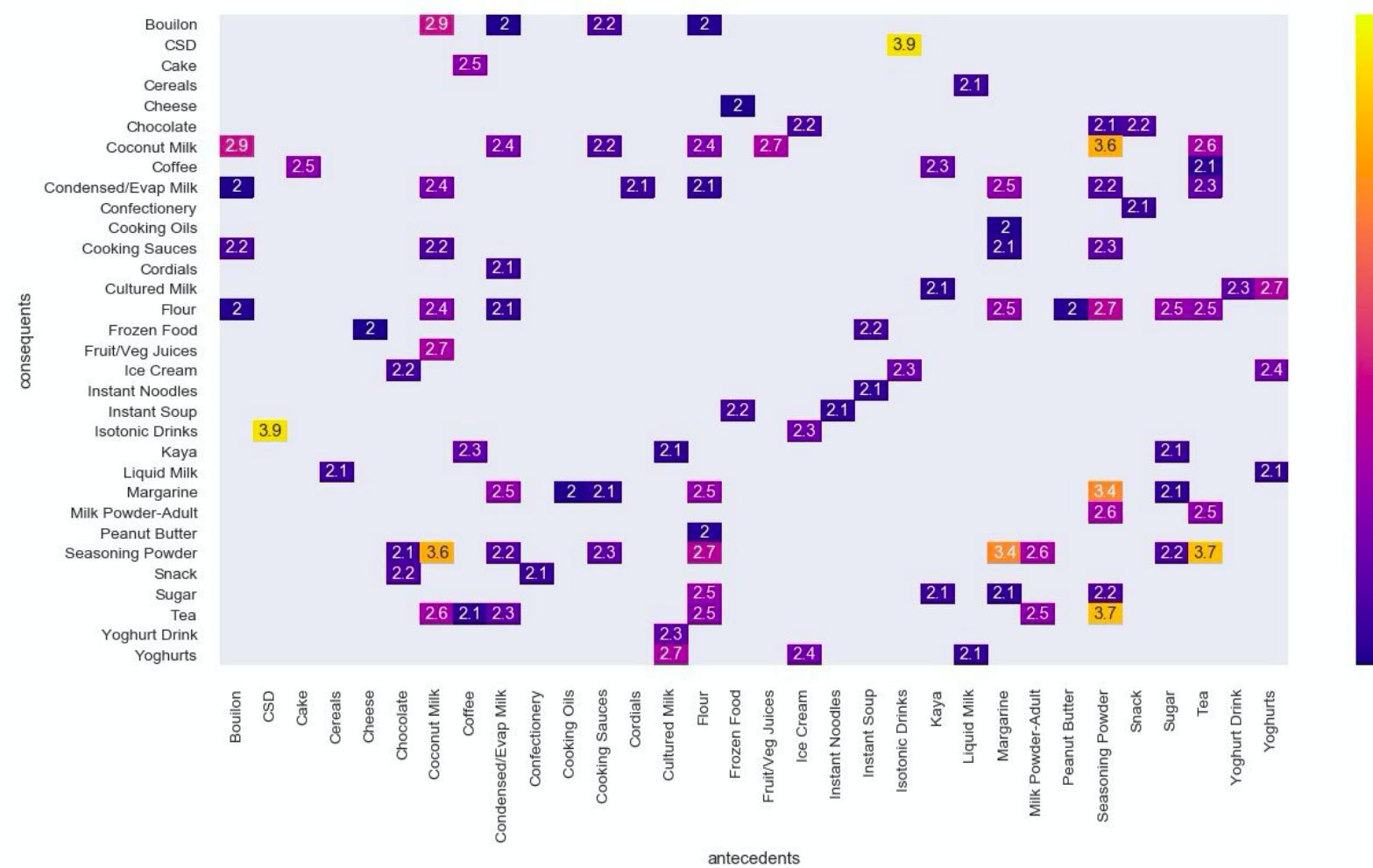
02

Approach

Analyse the market baskets of the 2 groups (RFM = 331 and RFM = 333) to understand their different purchasing patterns and obtain some insight as to how we can promote greater spending

MBA ANALYSIS

Lift of High Income and High Monetary Spending Consumers Pre-MCO Lift of High Income and High Monetary Spending Consumers Post-MCO



MBA ANALYSIS

Pre MCO		Post MCO	
Antecedents	Count	Antecedents	Count
Seasoning Powder	9	Coconut Milk	9
Flour	8	Flour	8
Coconut Milk	7	Sugar	7
Condensed/Evap Milk	7	Condensed/Evap Milk	6
Margarine	6	Seasoning Powder	6
Sugar	4	Bouillon	5

MBA - INSIGHTS

01

Lift for high spenders

General increase in lift between pre-MCO and post-MCO for High Spenders

02

Antecedent and Consequents pair

Increase in the number of antecedents and consequents pair that satisfies the minimum pre-set threshold of 2

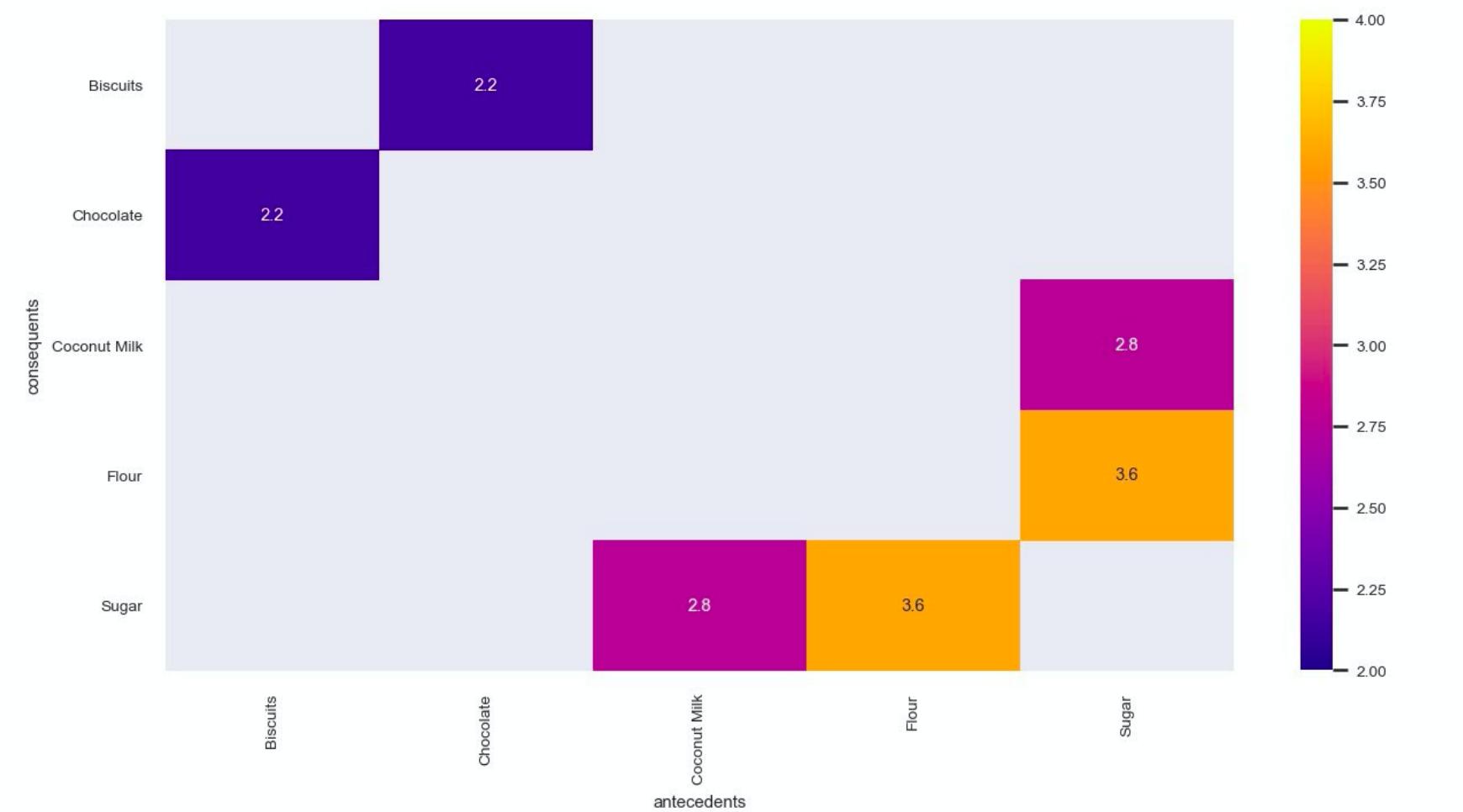
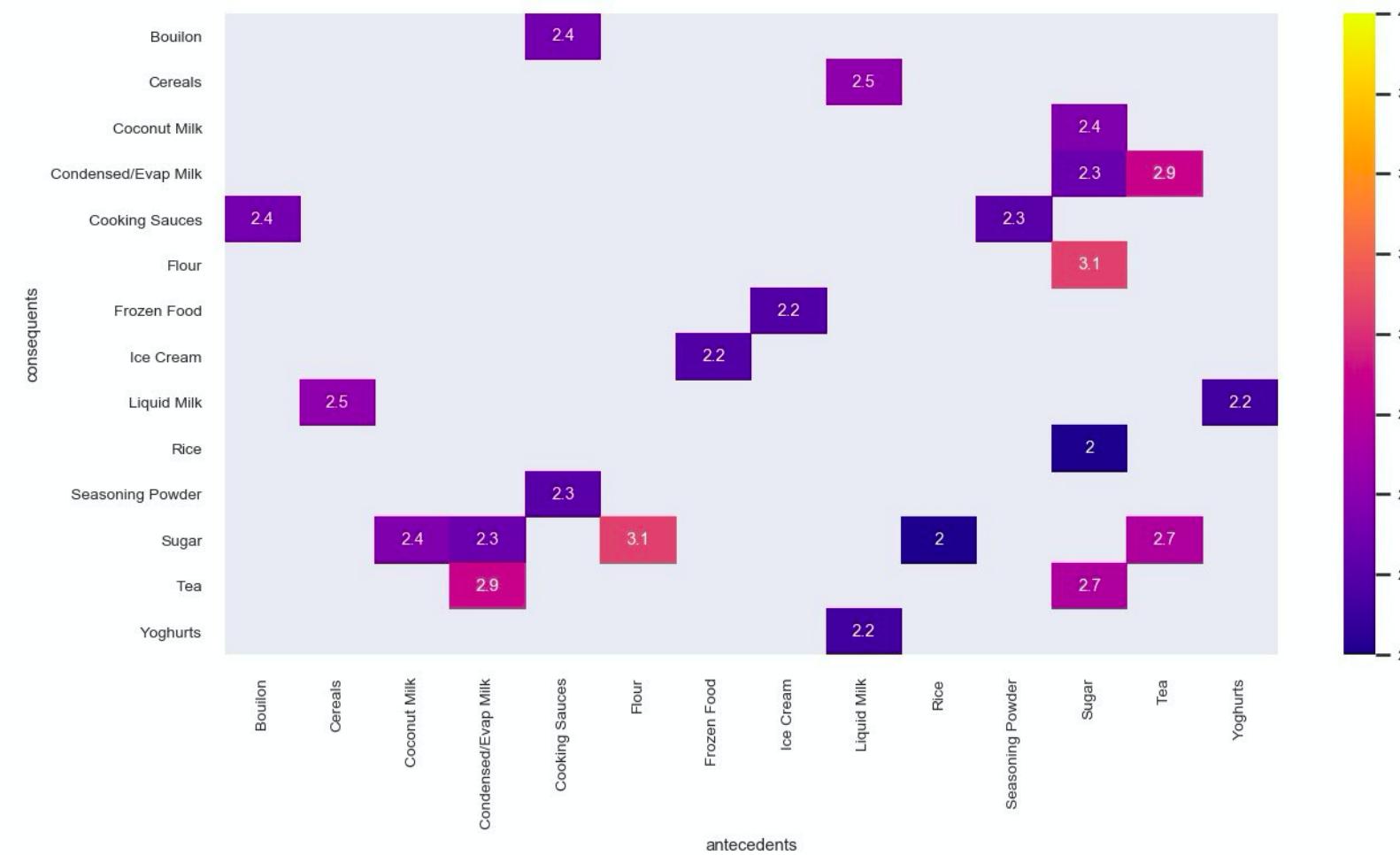
03

Customer Behaviours

Items like Coconut Milk and Sugar have experienced an increase in the number of Antecedent counts indicating that customers are purchasing a larger variety of goods in tandem with these items

MBA ANALYSIS

Lift of High Income and Low Monetary Spending Consumers Pre-MCO Lift of High Income and Low Monetary Spending Consumers Post-MCO



MBA ANALYSIS

Pre MCO		Post MCO	
Antecedents	Count	Antecedents	Count
Sugar	5	Sugar	2
Condensed/Evap Milk	2	Biscuits	1
Cooking Sauces	2	Chocolate	1
Liquid Milk	2	Coconut Milk	1
Tea	2	Flour	1

MBA - INSIGHTS

01

Antecedent and Consequents pair

Significant decrease in the number of Antecedents and Consequent pairs that satisfy the minimum pre set threshold of 2

02

Rearranging Sugar

Rearrange arrangements such that sugar is placed with items that are commonly purchased together, or even selling them as a bundle.

03

Customer Behaviours

Items like Sugar had a higher Antecedent count pre-MCO, indicating that customers are purchasing a smaller variety of goods in tandem with Sugar, post-MCO

Recommendation

Focus campaigning strategies towards customers in the central area and on items like sugar

Summary of insights and suggestions

Insight 1

Central area has highest proportion of affluent customers

Insight 2

Some affluent customers are not spending as much

Insight 3

High spenders buying more after MCO while low spenders buying less

Suggestion 1

Target more central area customers

Suggestion 2

Target 331 customers

Suggestion 3

Target items that can lead to higher sales



**Questions? Clarifications?
Let us know!**