Term 5 - ML3

Rupinder Kohli
March 2020 cohort

Items association rules at a grocery

Grocery dataset

Define Problem Statement	Perform EDA	Feature Engineering	Create Model	Evaluate Model

Define Problem & Approach to solve

Problem Statement:

As the owner of the store, I need to understand the products the customers are buying

- identify the fast moving products
- how do we place the products such that we are able to increase the overall sales

Approach:

To understand the fast moving products and applying the promotional rules for driving up sales, approach would be -

- find frequent associations between the items being bought
- recommend customers
- strategise product placements on the shelf
- apply promotions

The following algorithm can be applied to achieve the goal -

- Item-Item Recommendation
- Apriori

EDA

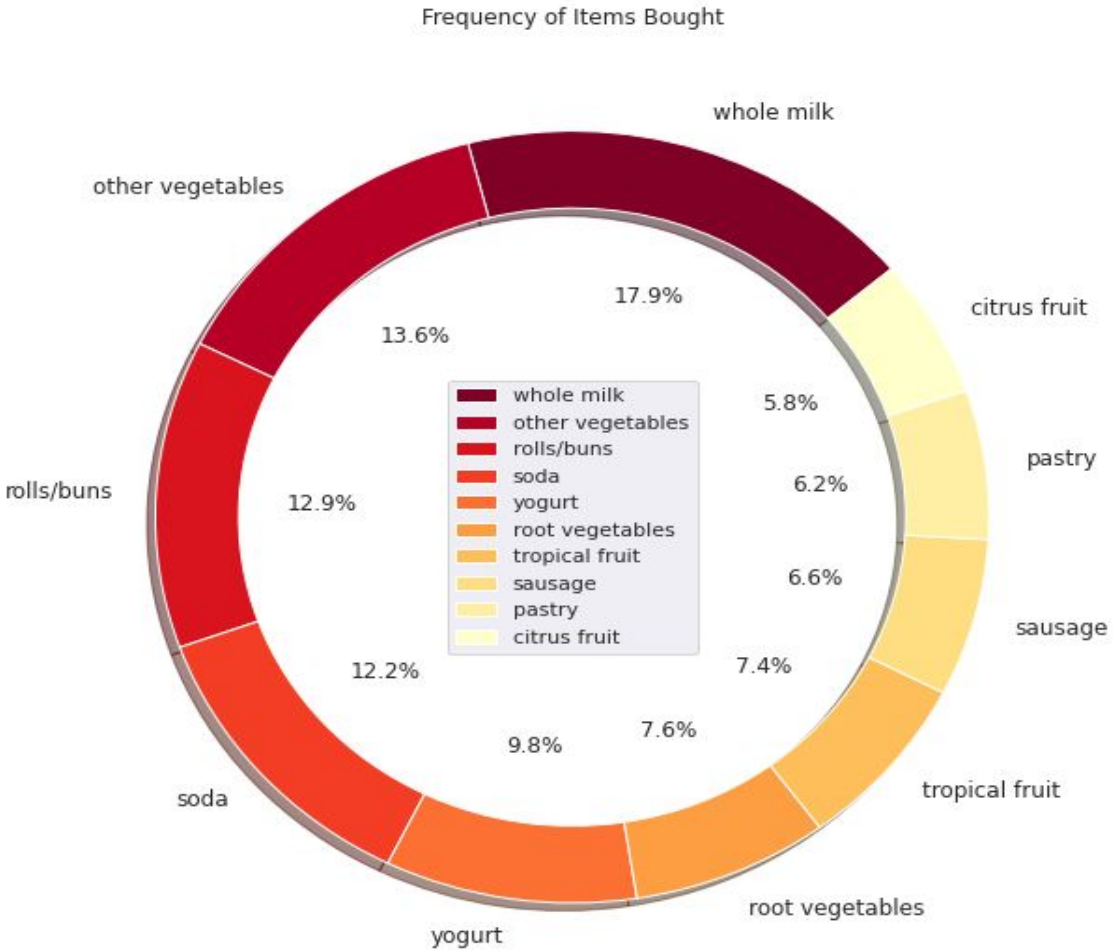
understanding the dataset

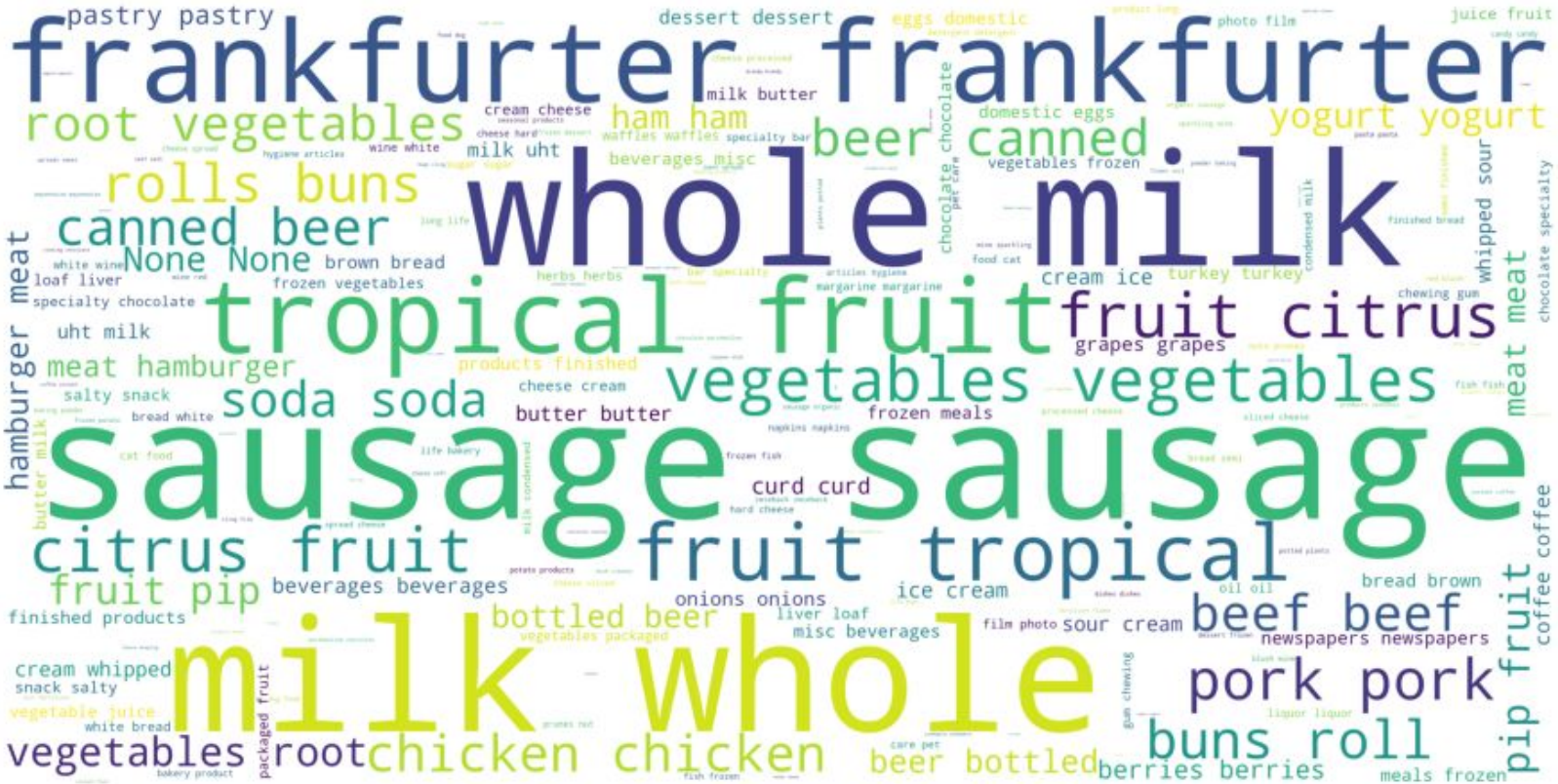
About Dataset

- Categorical dataset
- 9835 baskets
- 169 unique items
- Range of 1-32 items in baskets
- 463 repeated baskets
- Customers buy atleast 1 item from the store
- Atleast 30% baskets have atleast 6 items
- Less than 1% Baskets have more than 16 items

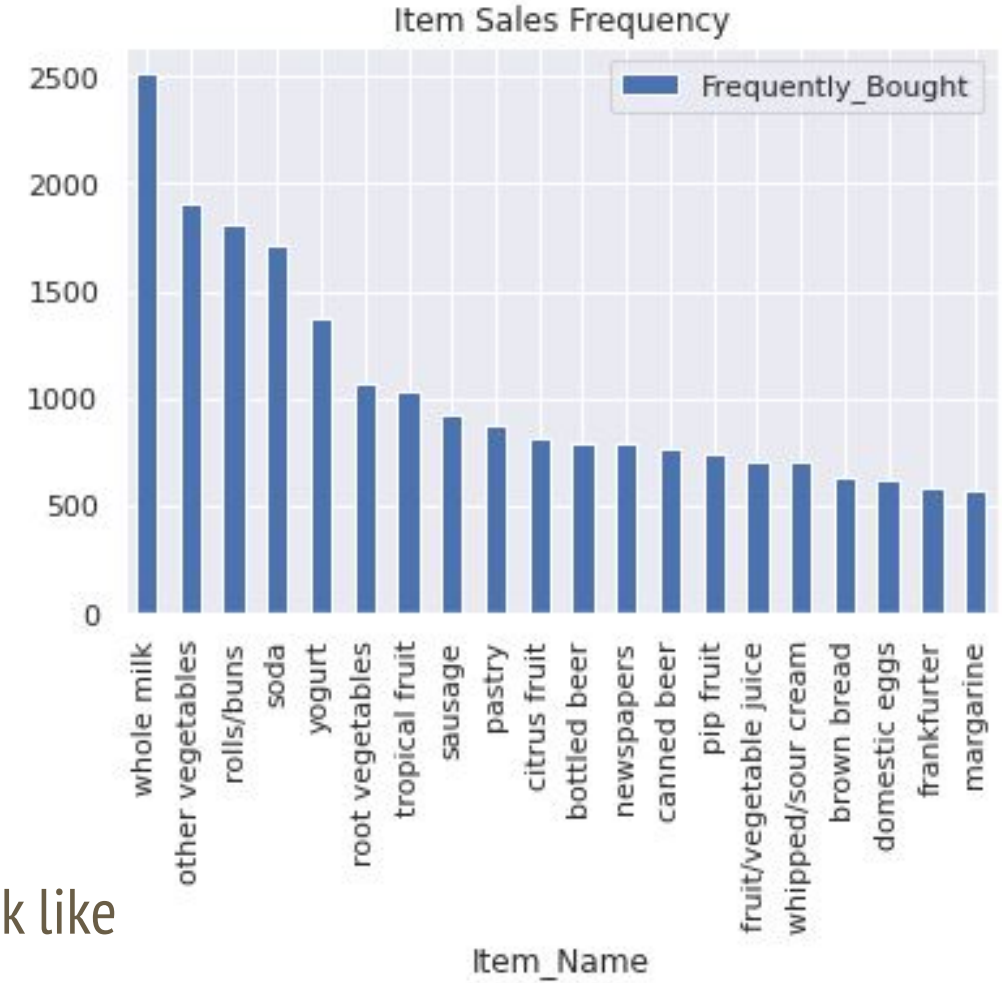
	Available	Availablity Percent	Total Nulls	Percent Nulls
item_count	9835	100.000000	0	0.000000
0	9835	100.000000	0	0.000000
1	7676	78.047789	2159	21.952211
2	6033	61.342145	3802	38.657855
3	4734	48.134215	5101	51.865785
4	3729	37.915608	6106	62.084392
5	2874	29.222166	6961	70.777834
6	2229	22.663955	7606	77.336045
7	1684	17.122522	8151	82.877478
8	1246	12.669039	8589	87.330961
9	896	9.110320	8939	90.889680
10	650	6.609049	9185	93.390951
11	468	4.758516	9367	95.241484
12	351	3.568887	9484	96.431113
13	273	2.775801	9562	97.224199
14	196	1.992883	9639	98.007117
15	141	1.433655	9694	98.566345
16	95	0.965938	9740	99.034062
17	66	0.671073	9769	99.328927

Frequently Bought Items





Fast going Items



How does the customer baskets look like

basket
citrus fruit,semi-finished bread,margarine,rea...
tropical fruit,yogurt,coffee
whole milk
pip fruit,yogurt,cream cheese,meat spreads
other vegetables,whole milk,condensed milk,lon...
whole milk,butter,yogurt,rice,abrasive cleaner
rolls/buns
other vegetables,uht-milk,rolls/buns,bottled b...
potted plants
whole milk,cereals

item_count	0	1	2	3	4
4	citrus fruit	semi-finished bread	margarine	ready soups	None
3	tropical fruit	yogurt	coffee	None	None
1	whole milk	None	None	None	None
4	pip fruit	yogurt	cream cheese	meat spreads	None
4	other vegetables	whole milk	condensed milk	long life bakery product	None
5	whole milk	butter	yogurt	rice	abrasive cleaner
1	rolls/buns	None	None	None	None
5	other vegetables	uht-milk	rolls/buns	bottled beer	liquor (appetizer)
1	potted plants	None	None	None	None
2	whole milk	cereals	None	None	None

0	1	2	3	4
29	132	88	118	-1
157	167	33	-1	-1
166	-1	-1	-1	-1
109	167	38	91	-1
102	166	34	85	-1
166	14	167	120	0
122	-1	-1	-1	-1
102	159	122	10	83
113	-1	-1	-1	-1
166	24	-1	-1	-1

RAW — ***ITEMISED*** — ***ENCODED***

Data Setup - Encode the data

Data Cleanup

- rename liqueur; liquor; liquor (appetizer) to liquor
- rename bags to cling film/bags
- Remove ‘shopping bags’ from the items list

Item-Item Recommendation System

Introduction - Item-Based Recommendation Engine

- Computes similarity between products based on baskets to suggest items that are bought along-with some other products.
- TfidfVectorizer function from scikit-learn, which transforms text to feature vectors that can be used as input to estimator.
- Further will be using the Cosine Similarity to calculate a numeric quantity that denotes the similarity between two baskets

Item-Based Recommendation Engine

- An item based recommendation system that computes similarity between products based on baskets.
- It will suggest items that are bought along-with some other products.

```
1 product_recommendations('popcorn')
```

```
['condensed milk',
 'coffee',
 'whole milk',
 'yogurt',
 'other vegetables',
 'pip fruit',
 'cream cheese',
 'semi-finished bread',
 'citrus fruit',
 'tropical fruit']
```

```
1 product_recommendations('salty snack')
```

```
['condensed milk',
 'coffee',
 'whole milk',
 'yogurt',
 'rolls/buns',
 'other vegetables',
 'pip fruit',
 'cream cheese',
 'uht-milk',
 'potted plants',
 'butter',
 'semi-finished bread',
 'citrus fruit',
 'tropical fruit']
```

Apriori Algorithm

Apriori algorithm was performed in 3 iterations

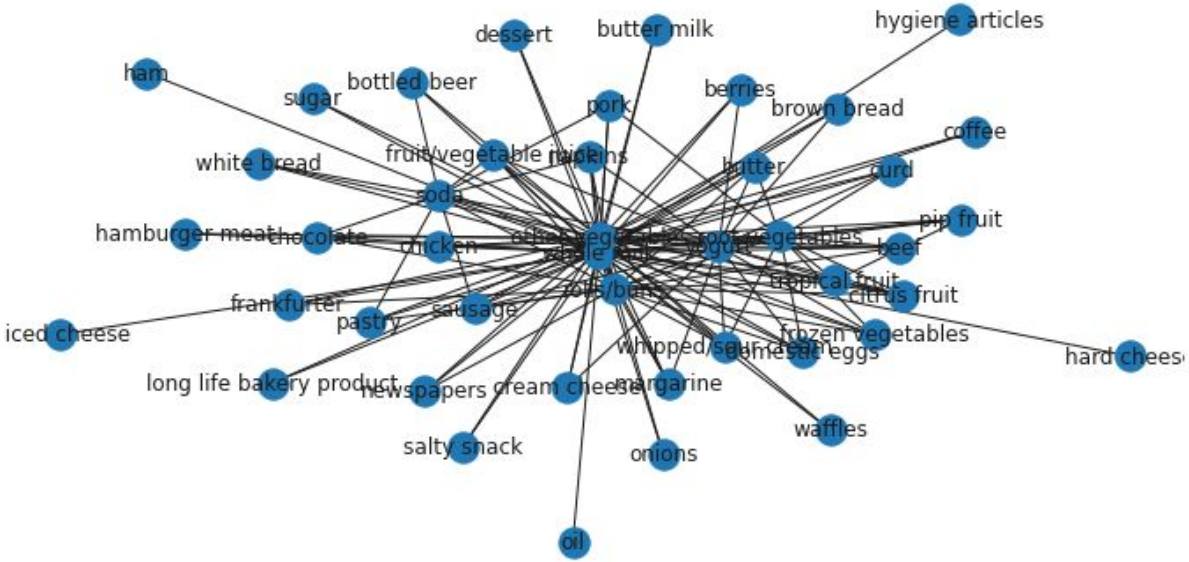
Applying Apriori

Approach will be to transform the available dataset such that

- each row represents a transaction and each column an individual item of the transaction
- determine the frequent item sets of this dataset
- the association rules can be determined
- setup the rule that if the SKU is part of atleast 3 baskets, it's classified as frequent item
- uses a "bottom up" approach, where frequent subsets are extended one item at a time

Iteration 1

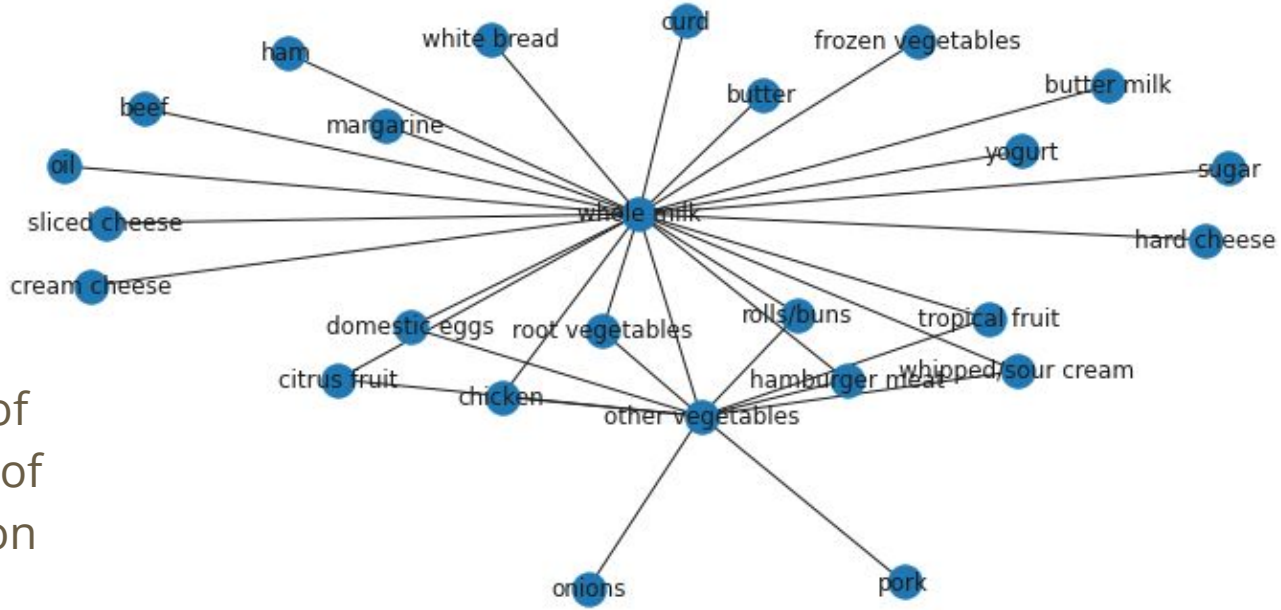
- Iteration 1 was performed with minimum support of 0.01 and threshold of 0.2 giving association rules of 224 items



Looking at the associations, to strategize sales and offers, next iteration with increased threshold should be performed

Iteration 2

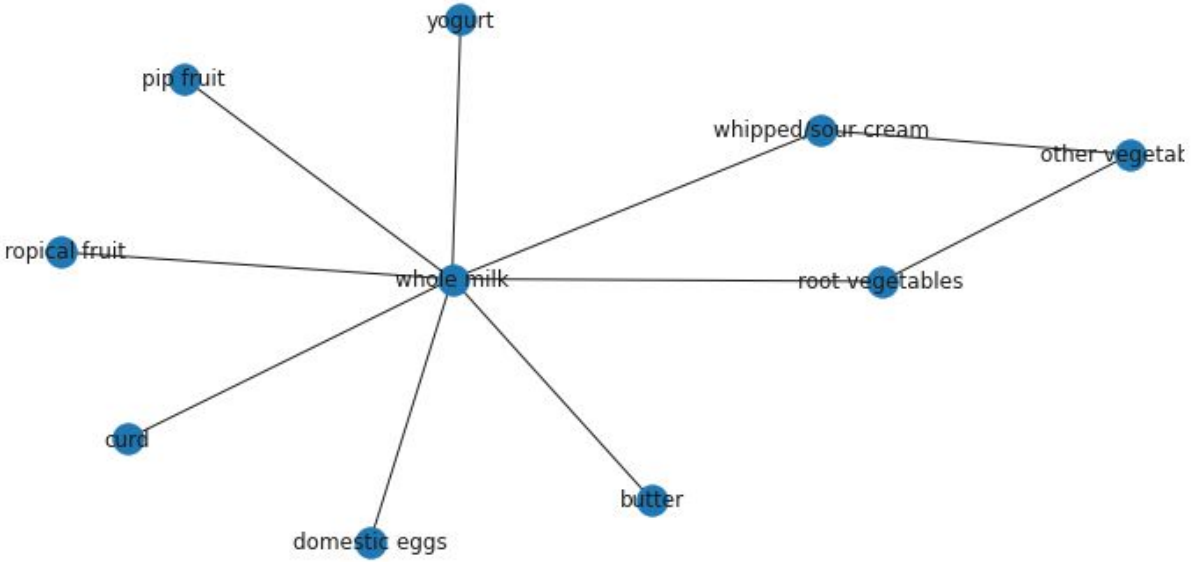
- Iteration 2 was performed with minimum support of 0.01 and threshold of 0.4 giving association rules of 61 items



The store could offer discounts on OIL for customer's buying X ltrs of milk to boost the sales of OIL

Iteration 3

- Iteration 3 was performed with minimum support of 0.25 and threshold of 0.39 giving association rules of 10 items



The store could offer customers for e.g. whipped cream when they are buying other vegetables or domestic eggs

Evaluation

Based on the generated Association Rules, the following parameters were observed -

- **Support - an indication of how frequently the itemset appears in the dataset.**
- **Confidence - an indication of how often the rule has been found to be true.**
- **Lift strengthens the fact that customer buying A will definitely buy B**

Per the 3 iterations performed, the most conservative was Iteration 3 (appended in the following slide)

Evaluation Metric

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(butter)	(whole milk)	0.055414	0.255516	0.027555	0.497248	1.946053	0.013395	1.480817
1	(curd)	(whole milk)	0.053279	0.255516	0.026131	0.490458	1.919481	0.012517	1.461085
2	(domestic eggs)	(whole milk)	0.063447	0.255516	0.029995	0.472756	1.850203	0.013783	1.412030
3	(root vegetables)	(other vegetables)	0.108998	0.193493	0.047382	0.434701	2.246605	0.026291	1.426693
4	(whipped/sour cream)	(other vegetables)	0.071683	0.193493	0.028876	0.402837	2.081924	0.015006	1.350565
5	(pip fruit)	(whole milk)	0.075648	0.255516	0.030097	0.397849	1.557043	0.010767	1.236375
6	(root vegetables)	(whole milk)	0.108998	0.255516	0.048907	0.448694	1.756031	0.021056	1.350401
7	(tropical fruit)	(whole milk)	0.104931	0.255516	0.042298	0.403101	1.577595	0.015486	1.247252
8	(whipped/sour cream)	(whole milk)	0.071683	0.255516	0.032232	0.449645	1.759754	0.013916	1.352735
9	(yogurt)	(whole milk)	0.139502	0.255516	0.056024	0.401603	1.571735	0.020379	1.244132

Conclusions

- Whole Milk and Other Vegetables are the bestseller of this grocery store
- Whole Milk shows association with 7 items -
  - yogurt, tropical fruit,other vegetables, root vegetables, domestic egg, Butter, whipped/sour cream.
- Recommending customers to buy Popcorn when buying any of the following
  - 'condensed milk', 'coffee', 'whole milk', 'yogurt', 'other vegetables', 'pip fruit', 'cream cheese', 'semi-finished bread','citrus fruit','tropical fruit'

Sales Strategy

Following strategies could be adopted to increase its sales considering the association we have seen between Whole Milk and its partners.

- Promotional offers/discounts can be applied by promoting a less sold product along with whole milk.
- Shelf strategy by placing vegetables and dairy items near to the ordering counter can be a good strategy to attract customers in buying all of these products and thereby increasing sales.
- Recommending
  - Whole-milk customers to buy whipped cream, popcorns
  - Whipped cream customers to buy other vegetables or domestic eggs

THANK YOU!