RETAIL ANALYTICS INDIVIDUAL ASSIGNMENT

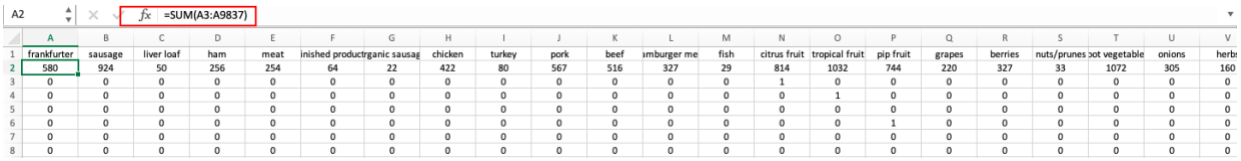
Summary:

For this task we were provided with the transaction dataset containing one month of real-world point-of-sale grocery store data, with 9835 transactions and 169 products. The values were represented in Binary with products as columns and transactions as rows; 1-entries correspond to the products included in a given transaction (shopping cart) and 0- entries correspond to the products not purchased in a given transaction.

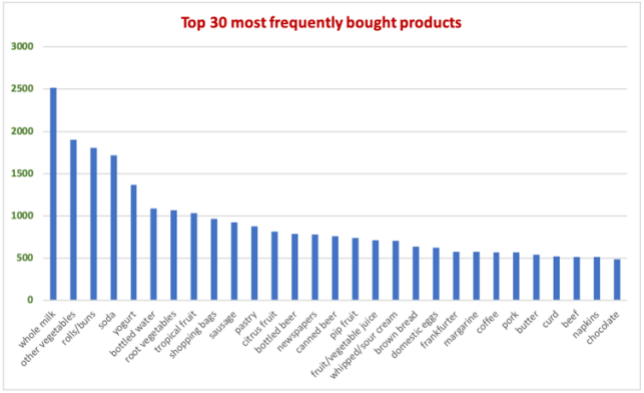
The objective is to infer most strongly associated product pairs (i.e. products frequently bought together) based on the provided shopping cart dataset. Also, to identify patterns and correlation in product purchasing to generate actionable insights that is to create potential interest for cross-selling and promotion of specific product combinations

Methodology:

Step 1: Finding the total number of times a product was bought out of 9835 transactions. This was done by using function SUM of all 1- entry binary transactions.



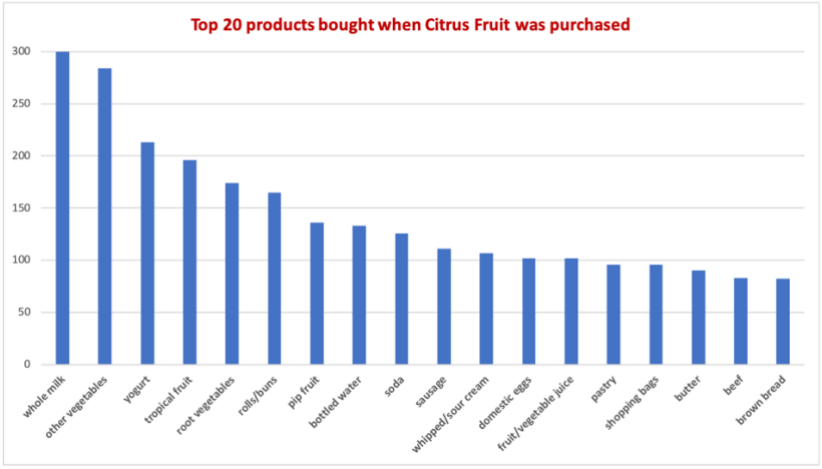
Step 2: Filtering the name of the products with total number of times the product of bought by the largest to smallest to find the top 20 frequently bought products.



Step 3: Identifying the top 12th most frequently bought product to analyze association rules.

Item	# of times bought	
whole milk	2513	
other vegetables	1903	
rolls/buns	1809	
soda	1715	
yogurt	1372	
bottled water	1087	
root vegetables	1072	
tropical fruit	1032	
shopping bags	969	
sausage	924	
pastry	875	
citrus fruit	814	Top 12th most frequently bought product
bottled beer	792	

Step 4: Filtering the transactions to only show the ones where citrus fruit has binary value of 1 (citrus fruit was bought). This will display the top other frequently bought items when citrus fruit was purchased.



Step 5: Calculating key metrics such as support, confidence and lift values to analyze association rules.

Product	frankfurter	sausage	liver loaf	ham	meat	finished products	organic sausage	chicken	turkey	pork	beef	hamburger meat	fish	citrus fruit
	580	924	50	256	254	64	22	422	80	567	516	327	29	814
NAB	frankfurter	sausage	liver loaf	ham	meat	finished products	organic sausage	chicken	turkey	pork	beef	hamburger meat	fish	citrus fruit
	64	111	5	28	34	7	4	68	17	64	83	39	4	814
Number of product transactions	9835													
Support	=B2/$B$7	9.4%	0.5%	2.6%	2.6%	0.7%	0.2%	4.3%	0.8%	5.8%	5.2%	3.3%	0.3%	8.3%

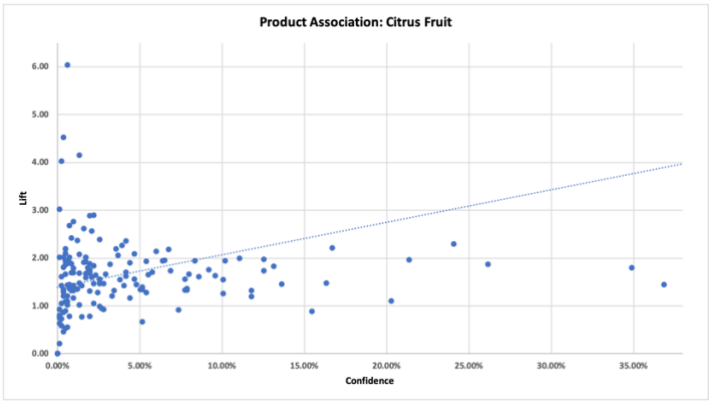
NAB	frankfurter	sausage	liver loaf	ham	meat	finished products	organic sausage	chicken	turkey	pork	beef	hamburger meat	fish	citrus fruit
	64	111	5	28	34	7	4	68	17	64	83	39	4	814
Number of transactions NAB	814													
Confidence	=B5/$B$10	13.64%	0.61%	3.44%	4.18%	0.86%	0.49%	8.35%	2.09%	7.86%	10.20%	4.79%	0.49%	100.00%

Support	5.9%	9.4%	0.5%	2.6%	2.6%	0.7%	0.2%	4.3%	0.8%	5.8%	5.2%	3.3%	0.3%	8.3%
Number of transactions NAB	814													
Confidence	7.86%	13.64%	0.61%	3.44%	4.18%	0.86%	0.49%	8.35%	2.09%	7.86%	10.20%	4.79%	0.49%	100.00%
Lift	=B11/B8	1.45	1.21	1.32	1.62	1.32	2.20	1.95	2.57	1.36	1.94	1.44	1.67	12.08

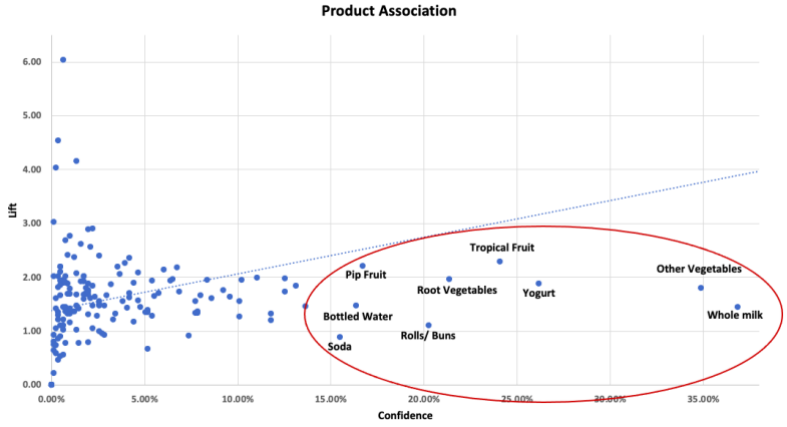
Step 6: Filtering the key metrics by confidence largest to smallest and ten by lift values largest to smallest to analyze association rules.

Product	Support	Confidence	Lift
citrus fruit	8.3%	100.00%	12.08
whole milk	25.6%	36.86%	1.44
other vegetables	19.3%	34.89%	1.80
yogurt	14.0%	26.17%	1.88
tropical fruit	10.5%	24.08%	2.29
root vegetables	10.9%	21.38%	1.96
rolls/buns	18.4%	20.27%	1.10
pip fruit	7.6%	16.71%	2.21
bottled water	11.1%	16.34%	1.48
soda	17.4%	15.48%	0.89
sausage	9.4%	13.64%	1.45
whipped/sour cream	7.2%	13.14%	1.83
domestic eggs	6.3%	12.53%	1.97
fruit/vegetable juice	7.2%	12.53%	1.73
pastry	8.9%	11.79%	1.33
shopping bags	9.9%	11.79%	1.20
butter	5.5%	11.06%	2.00
beef	5.2%	10.20%	1.94
brown bread	6.5%	10.07%	1.55

Step 7: Plotting a graph between Lift and confidence to find the most promising products based on association rules for the product citrus fruit.



Step 8: Identifying the top 10 most promising products based on lift and confidence values for the product citrus fruit.



Analysis and Solution:

By looking at the graph, we can see that the top products bought with citrus fruits include whole milk, other vegetables, yogurt, tropical fruit, root vegetables, rolls and buns, pip fruits, bottled water and soda. We can say that since the fruits and vegetables are normally displayed in one section this has led to possibly buying both the products together. While we can assume that products such as milk, yogurt and rolls/ buns were bought with addition to citrus fruits due to the close proximity of freezer section to the vegetables while also being great breakfast options. As we know the citrus fruits are very popular to add in water and soda to drink in summers, we can assume that this has reflected on both the products being bought together.

We should display the top products in close proximity to each other so consumers can spot them easily and purchase. Display complimentary products beside each product to influence customer behavior. Show organic, premium or high-end version of the products to influence consumer decisions. We should create promotional, combo or bundling offers of citrus fruits with the top 10 promising products where customers can get a discount if they buy the products together thus increasing sales.

References:

Giannini, B., & Paramonov, P. (2022, July 1). *Course task: Shopping cart analysis*. Hult. https://mycourses.hult.edu/courses/3349558/files/221491204?module_item_id=76249145&fd_cookie_set=1