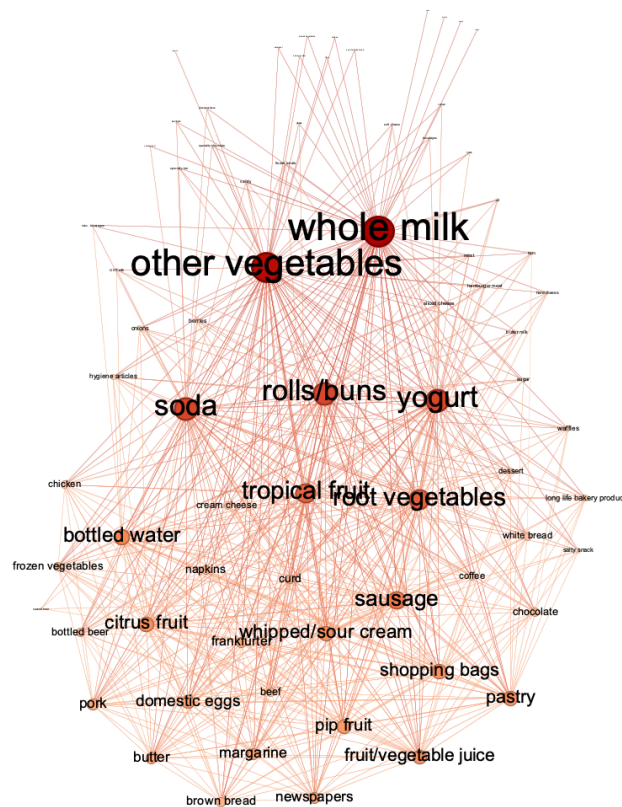# Market Basket Analysis

The graph below shows the association between all items with a minimum support of .005. There was no minimum threshold for lift and confidence, so all association rules, regardless of their strength or significance are shown. The larger the node in this graph, the higher the degree.


Let's take a look at the association rules with the highest support, which is a measure for the proportion of transactions that contain all items in the antecedent and consequent.

antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(other vegetables)	(whole milk)	0.193493	0.255516	0.074835	0.386758	1.513634
(whole milk)	(other vegetables)	0.255516	0.193493	0.074835	0.292877	1.513634
(rolls/buns)	(whole milk)	0.183935	0.255516	0.056634	0.307905	1.205032
(whole milk)	(rolls/buns)	0.255516	0.183935	0.056634	0.221647	1.205032
(yogurt)	(whole milk)	0.139502	0.255516	0.056024	0.401603	1.571735
(whole milk)	(yogurt)	0.255516	0.139502	0.056024	0.219260	1.571735
(root vegetables)	(whole milk)	0.108998	0.255516	0.048907	0.448694	1.756031
(whole milk)	(root vegetables)	0.255516	0.108998	0.048907	0.191405	1.756031
(root vegetables)	(other vegetables)	0.108998	0.193493	0.047382	0.434701	2.246605
(other vegetables)	(root vegetables)	0.193493	0.108998	0.047382	0.244877	2.246605

Viewing these item pairs, we see that 4/10 of the top 10 item sets contain milk as the antecedent. Although we see that these pairs appear quite often in transactions, the confidence values for each of these pairs is below .5, meaning that for all of these pairs, if milk is present then there is less than a 50% chance that B is present.

When you think about how people grocery shop, this makes a lot of sense. Whole milk is one of the most commonly purchased ingredients at the grocery store, so it appears in a lot of transactions. This is backed up by the plot below, which shows the items that appeared in the most transactions, as suspected, whole milk appeared the most.
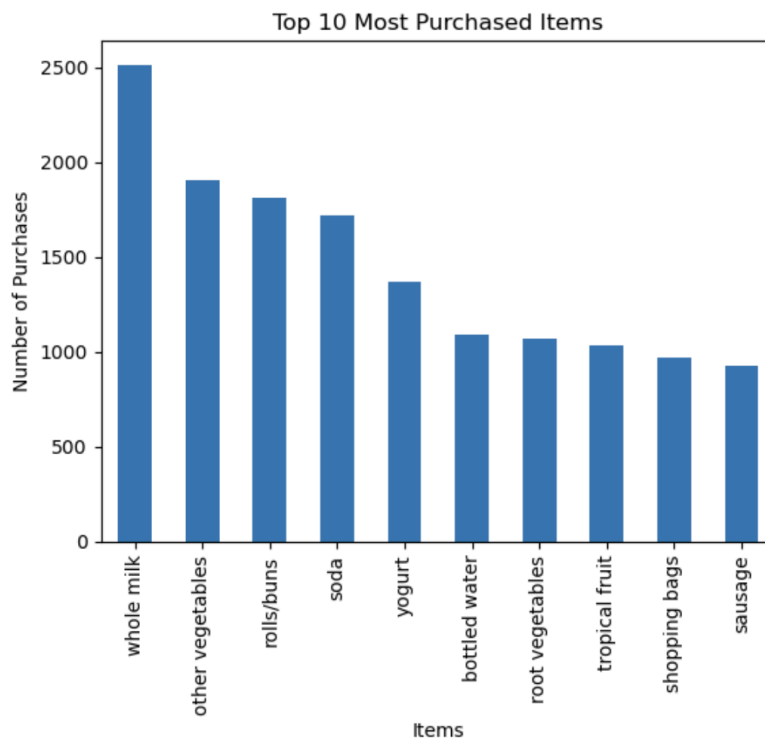
As a result of this graph and our confidence values, we can conclude that the presence of whole milk in our basket is not a great predictor for the presence of other vegetables, rolls/buns,

yogurt, or root vegetables, rather these items just happen to be purchased quite often. Stores can leverage the popularity of these items by placing emphasis on product placement in stores.

To create a set of rules with stronger associations, we created subsets of rules based on some confidence and lift thresholds.
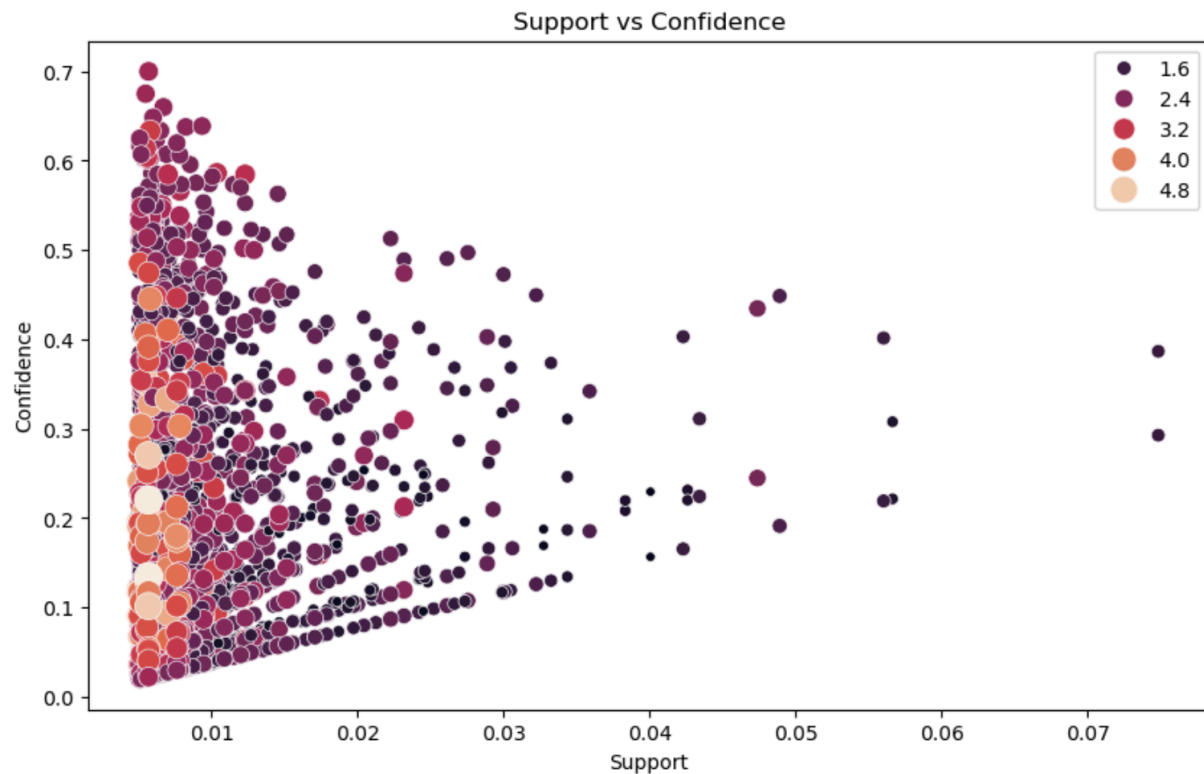

Based on the graph above, we set our confidence threshold at .5, meaning we only want to consider association rules where if the antecedent is present, there is a greater than 50% chance that the consequent is present. The table below lists the top 10 association rules by confidence.
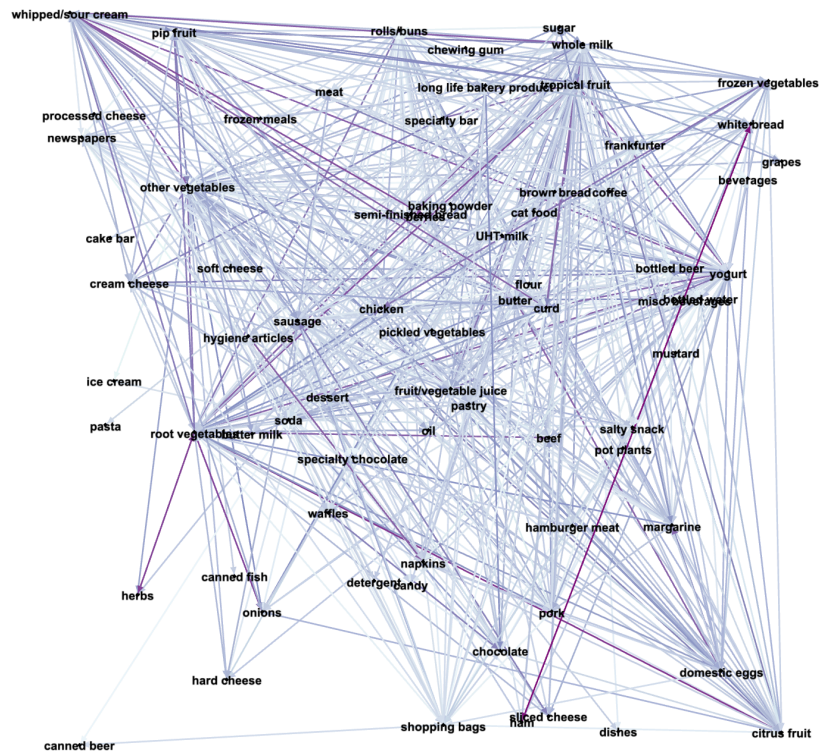
Association Rules with Confidence of at Least .5:
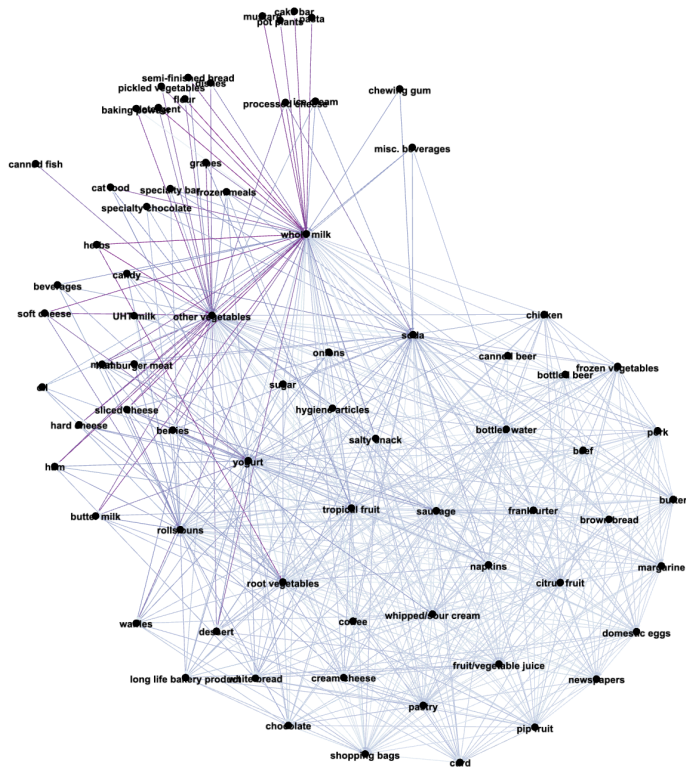
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
2938	(tropical fruit, yogurt, root vegetables)	(whole milk)	0.008134	0.255516	0.005694	0.700000	2.739554
2812	(other vegetables, root vegetables, pip fruit)	(whole milk)	0.008134	0.255516	0.005491	0.675000	2.641713
1520	(whipped/sour cream, butter)	(whole milk)	0.010168	0.255516	0.006711	0.660000	2.583008
2514	(whipped/sour cream, pip fruit)	(whole milk)	0.009253	0.255516	0.005999	0.648352	2.537421
1526	(yogurt, butter)	(whole milk)	0.014642	0.255516	0.009354	0.638889	2.500387
1508	(root vegetables, butter)	(whole milk)	0.012913	0.255516	0.008236	0.637795	2.496107
1740	(tropical fruit, curd)	(whole milk)	0.010269	0.255516	0.006507	0.633663	2.479936
2787	(root vegetables, citrus fruit, whole milk)	(other vegetables)	0.009151	0.193493	0.005796	0.633333	3.273165
2826	(other vegetables, yogurt, pip fruit)	(whole milk)	0.008134	0.255516	0.005084	0.625000	2.446031
1806	(domestic eggs, pip fruit)	(whole milk)	0.008643	0.255516	0.005389	0.623529	2.440275

We see that 9 of the top 10 association rules, according to confidence, have whole milk as a consequent. This confirms that whole milk is a staple item in many transactions, suggesting that purchasing whole milk may be a driving factor of customers visiting the store. Given its purchase frequency, it will be important for the store to make sure that they have enough inventory of whole milk.

Support vs Lift
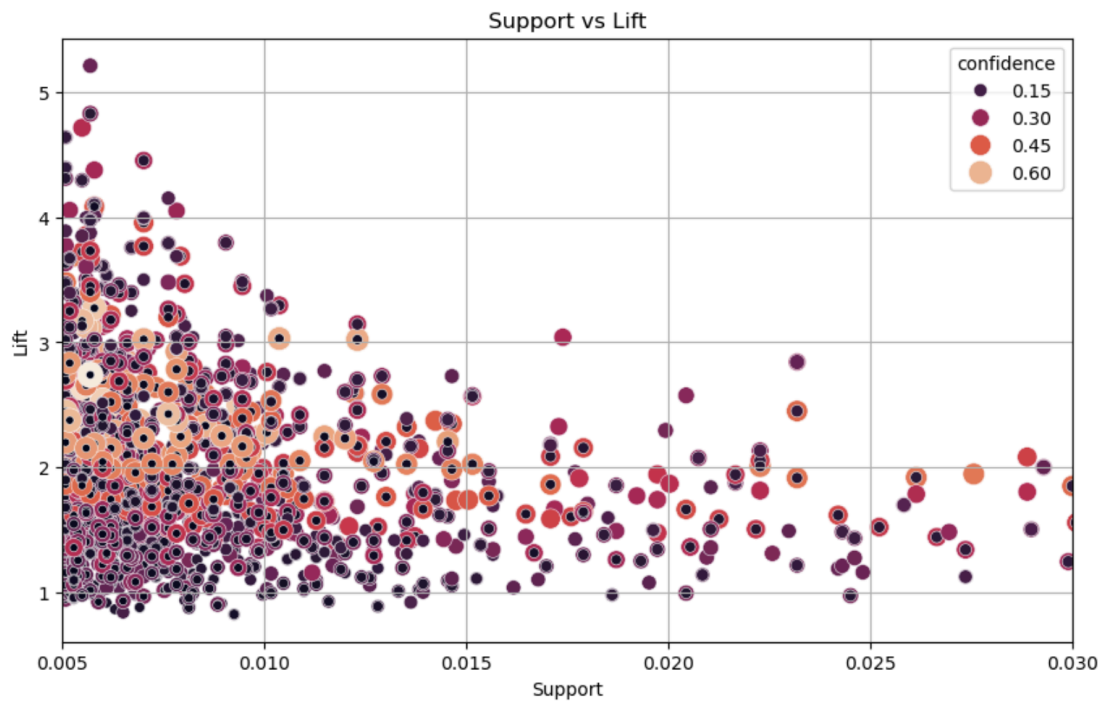
Based on the figure above and using our understanding of lift, we set the lift threshold to be any association rules where the lift is greater than one. This means that the antecedent and consequent are positively associated.

Association Rules with Lift of at Least 1:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
2944	(tropical fruit, whole milk)	(root vegetables, yogurt)	0.042298	0.025826	0.005694	0.134615	5.212371
2945	(root vegetables, yogurt)	(tropical fruit, whole milk)	0.025826	0.042298	0.005694	0.220472	5.212371
2942	(tropical fruit, root vegetables)	(yogurt, whole milk)	0.021047	0.056024	0.005694	0.270531	4.828814
2947	(yogurt, whole milk)	(tropical fruit, root vegetables)	0.056024	0.021047	0.005694	0.101633	4.828814
2819	(root vegetables, pip fruit)	(other vegetables, whole milk)	0.015557	0.074835	0.005491	0.352941	4.716272
2818	(other vegetables, whole milk)	(root vegetables, pip fruit)	0.074835	0.015557	0.005491	0.073370	4.716272
694	(ham)	(white bread)	0.026029	0.042095	0.005084	0.195312	4.639851
695	(white bread)	(ham)	0.042095	0.026029	0.005084	0.120773	4.639851
2873	(tropical fruit, root vegetables)	(other vegetables, whole milk)	0.021047	0.074835	0.007016	0.333333	4.454257
2876	(other vegetables, whole milk)	(tropical fruit, root vegetables)	0.074835	0.021047	0.007016	0.093750	4.454257

One interesting thing to note from this table is the presence of both ham -> white bread rule and the white bread -> ham rule. If we had to guess, this has to do with what ham and white bread are used for, sandwiches. It is interesting to note that the lift is high for both of these rules; however, the confidence is low.

What does this mean for the store? High lift and low confidence means that this combination is popular among a specific customer base, but this combination is not widespread across the entire customer base. This presents an opportunity for cross-promotion because we

know that these items are strongly associated but not regularly bought together. Offering a promotion on purchasing both items, may encourage a larger set of customers to purchase ham and white bread together.

Overall, association rules provide valuable insights, but they should be interpreted in the context of store strategy and the behavior of consumers to effectively guide inventory management and marketing efforts.