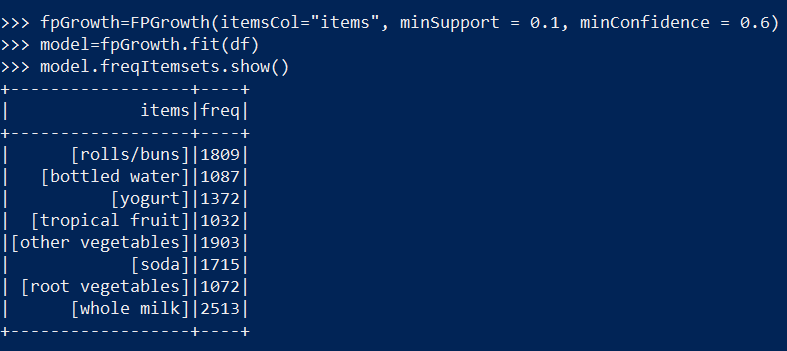
Frequent Pattern/Association Rule Mining with Apache Spark; PySpark codes running in PySpark shells. I have performed co-occurrence analysis on a marginably sized dataset, to demonstrate a models ability indicative to consumer purchase probability. Co-occurrence analysis is useful for store layouts, identifying healthcare patients, and marketing.

Our frequent pattern mining will utilize the purchase data collected from one month of operation at a real-world grocery store. The data contains 9,835 transactions or about 327 transactions per day (roughly 30 transactions per hour in a 12-hour business day), suggesting that the retailer is not particularly large, nor is it particularly small.

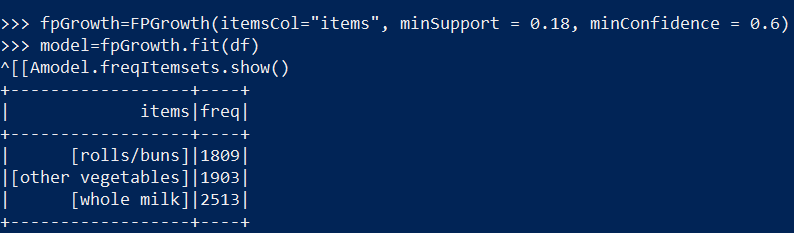
min\_support, max\_support, and confidence values are used to find and predict association rules.

Q1.



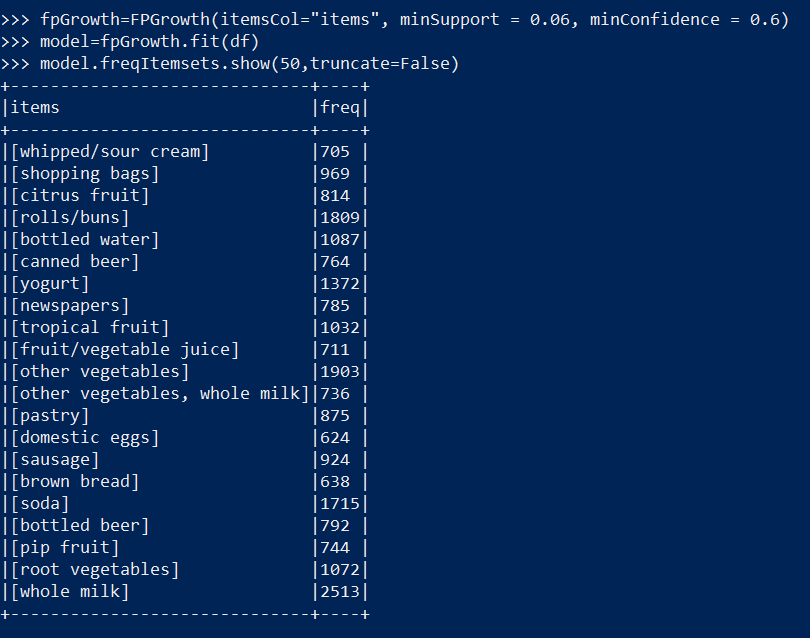
**Answer: 1st: Whole Milk, 2nd: Other Vegetables, and 3rd: Rolls/Buns.**

Q2.



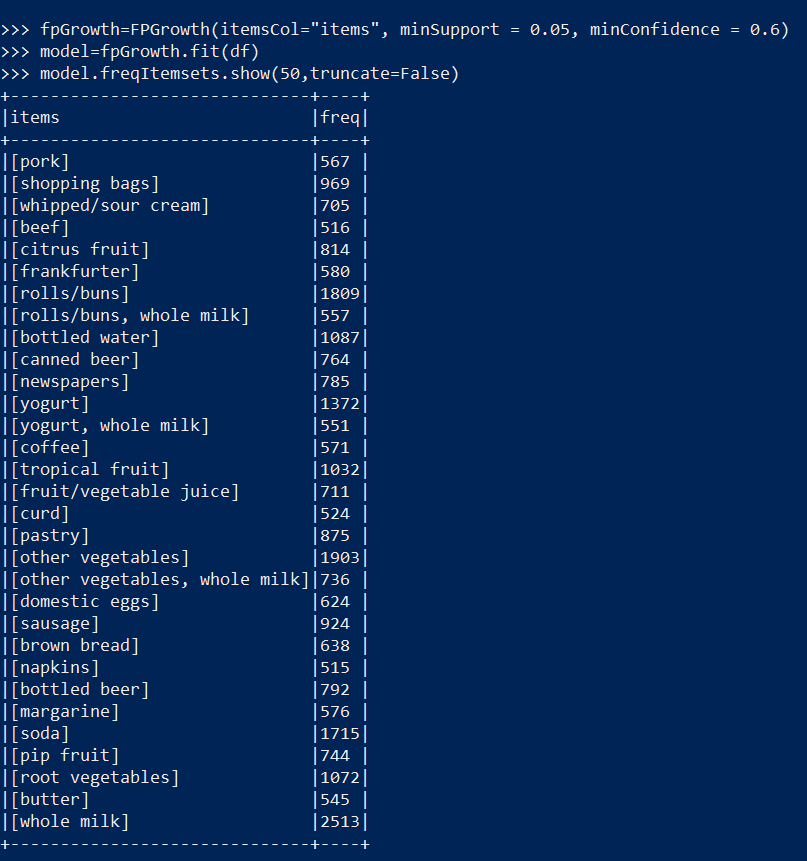
**Answer: The support for the three itemsets is 0.18 or 18%.**

Q3.



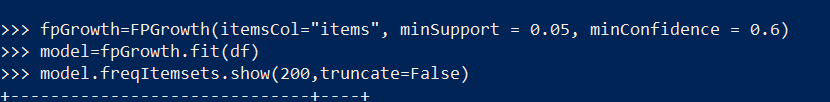
**Answer: [Other Vegetables, Whole Milk] = 736**

Q4.



**Answer: [other vegetables, whole milk] : 736, [yogurt, whole milk] : 551, [rolls/buns, whole milk]: 557.**

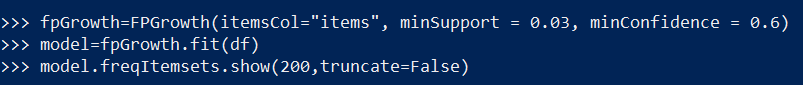
**We get more itemsets when the minSupport is lower. 0.06 support or 6% gives one itemset of size 2. 0.05 or 5% gives us three itemsets of size 2.**

Q5. 

Answer: **Rolls/buns is most frequently purchased with:**

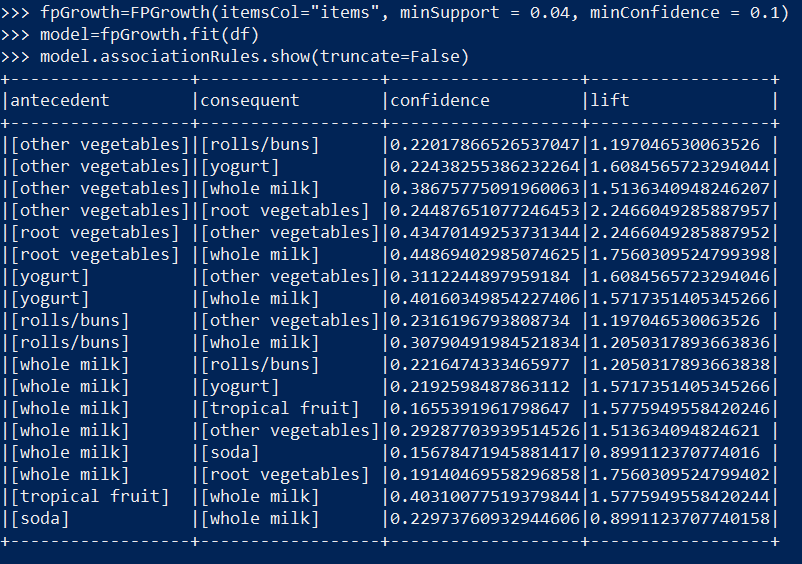


Q6.





Q7.



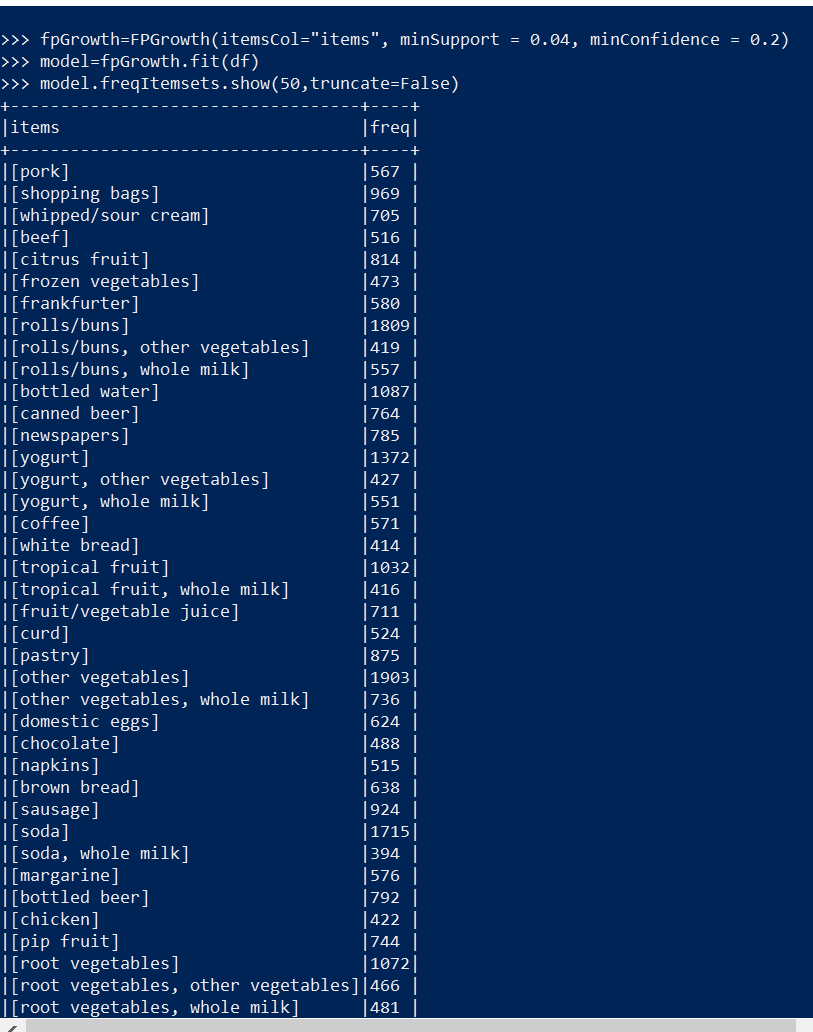
Answer: Confidence is 0.38675775091960063.

Q8:

(from screenshot above)

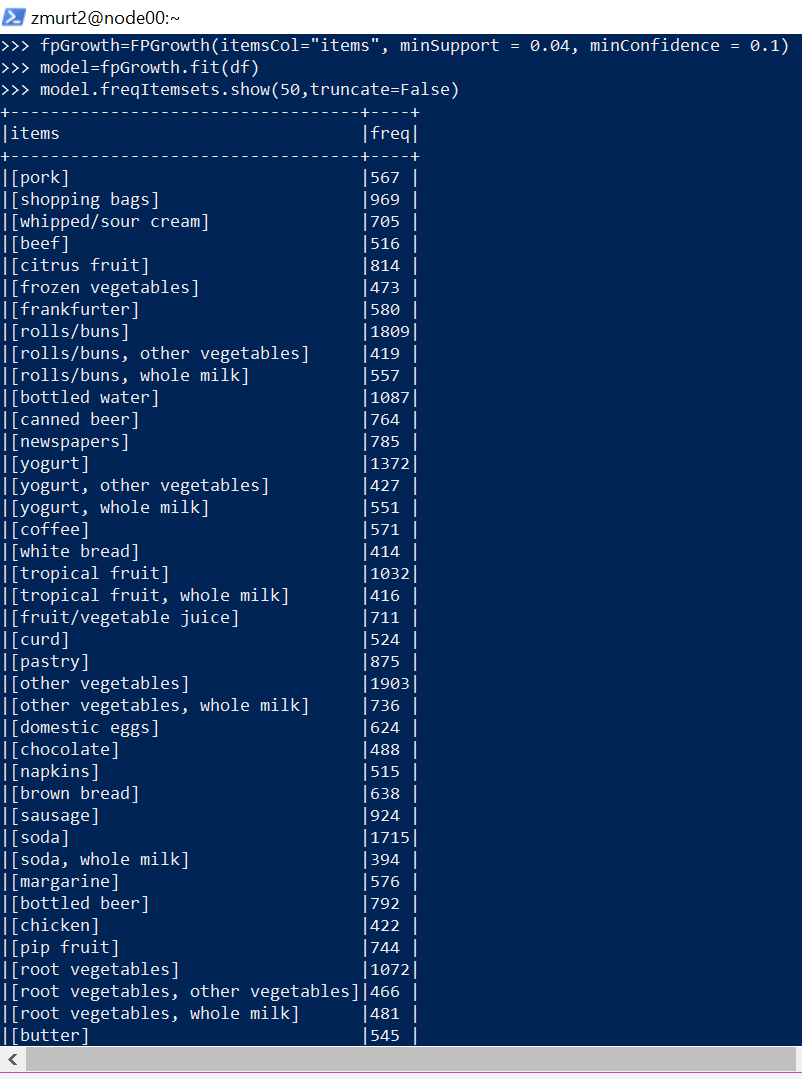
Confidence is: 0.40160349854227406

Q9.



**Answer: rolls/buns, yogurt, tropical fruit, other vegetables, root vegetables, and soda might be purchased with whole milk.**

**Q10.**



**rolls/buns, yogurt, tropical fruit, other vegetables, root vegetables, and soda might be purchased with whole milk.**