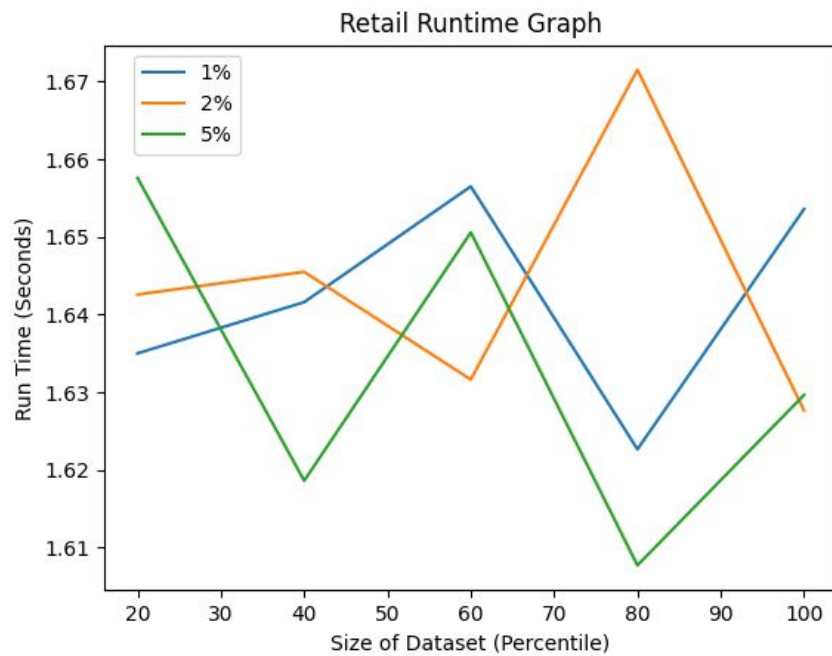


Midterm Lab Report

CSCI 4030u

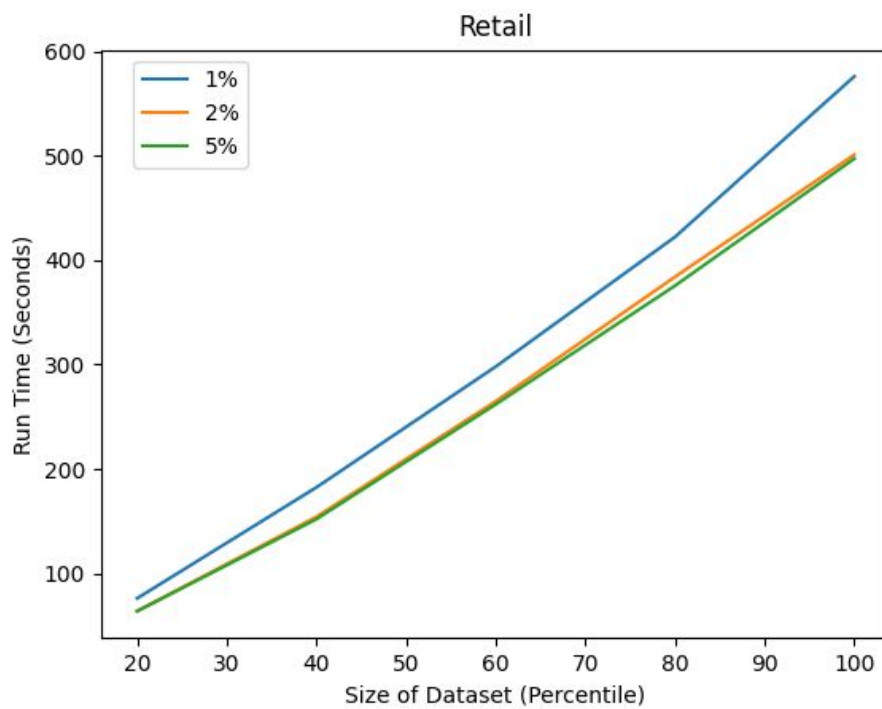
Andrew Murdoch
Andre Forbes

Lab 1:



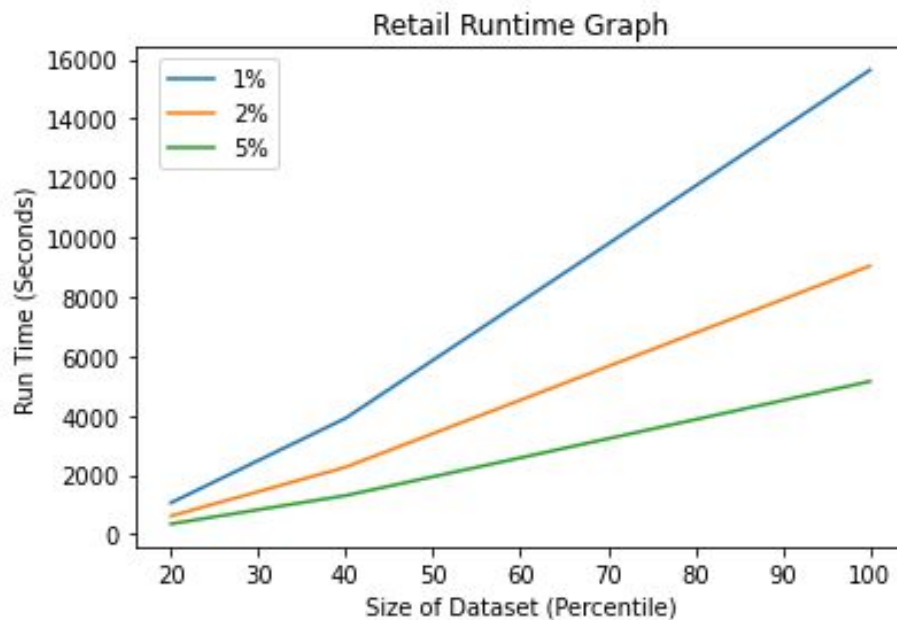
The Apriori implementation by the Apyori library runs the retail dataset very well at various minimum support levels.

Lab 2:



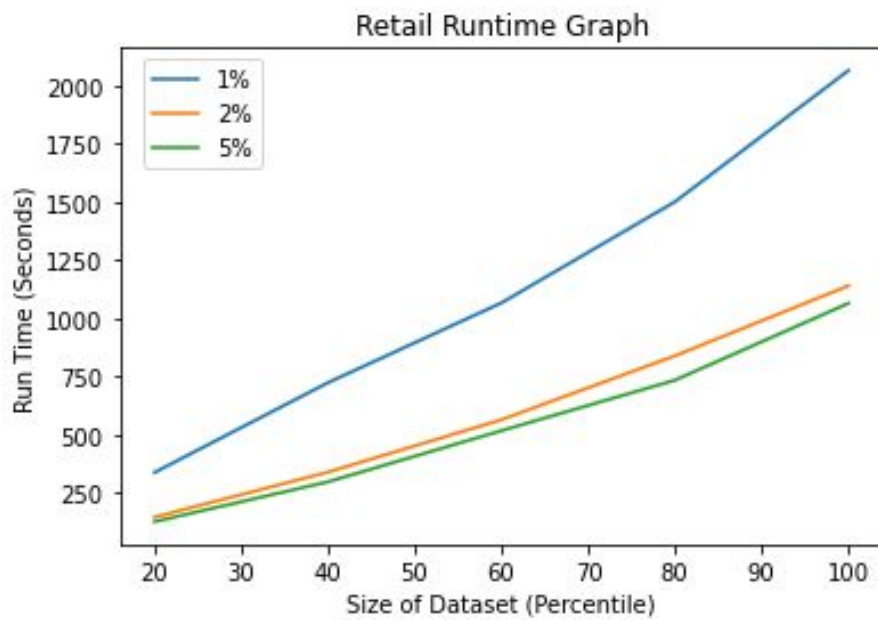
Lab 3:

For the sake of running times, the 60% , 80% and 100% tests were all estimated by multiplying the 20% values by 3.7n to approximate the values. Where 'n' is the test that took place (40% is 1, 60% is 2, etc.).



Comparing PCY to Apriori, it is evident that Apriori has a much more efficient run time compared to the PCY algorithm. This is the trade off when constructing an algorithm for memory efficiency.

Lab 4:



False positives singles:

	1%	2%	5%
20%	246	71	19
40%	232	73	11
60%	222	70	11
80%	232	69	11
100%	232	74	17

False positives doubles:

	1%	2%	5%
20%	282	75	10
40%	285	76	11
60%	255	61	13
80%	244	59	14
100%	280	65	12

False positives triples:

	1%	2%	5%
20%	134	37	4
40%	124	36	4
60%	113	34	5
80%	121	29	5
100%	125	29	5

When running this dataset with the various algorithms, the most effective way to get the results is to use the built in Apriori algorithm in python as it returns the results within the shortest amount of time.

Lab 5:

For the sake of running times, the 60% , 80% and 100% tests were all estimated by multiplying the 20% values by 3.8n to approximate the values. Where 'n' is the test that took place (40% is 1, 60% is 2, etc.).

